

Séquences d'explication lexicale dans l'enseignement du français par visioconférence : une approche multimodale

Benjamin Thomas Holt

▶ **To cite this version:**

Benjamin Thomas Holt. Séquences d'explication lexicale dans l'enseignement du français par visioconférence : une approche multimodale. Linguistique. Université de Lyon, 2018. Français. ⟨NNT : 2018LYSE2066⟩. ⟨tel-01976827⟩

HAL Id: tel-01976827

https://theses.hal.science/tel-01976827v1

Submitted on 10 Jan 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.



HAL Authorization

N° d'ordre NNT : 2018LYSE2066

THESE de DOCTORAT DE L'UNIVERSITÉ DE LYON

Opérée au sein de

L'UNIVERSITÉ LUMIÈRE LYON 2

École Doctorale : ED 484 Lettres, Langues, Linguistique et Arts

Discipline : Sciences du langage

Soutenue publiquement le 12 octobre 2018, par :

Benjamin HOLT

Séquences d'explication lexicale dans l'enseignement du français par visioconférence : une approche multimodale.

Devant le jury composé de :

Heather HILTON, Professeure des universités, Université Lumière Lyon 2, Présidente

Lucile CADET, Professeure des universités, Université de Cergy-Pontoise, Rapporteure

Annick RIVENS MOMPEAN, Professeure des universités, Université Lille III, Rapporteure

Thierry SOUBRIÉ, Maître de conférences, Université Grenoble Alpes, Examinateur

Nicolas GUICHON, Professeur des universités, Université Lumière Lyon 2, Co-Directeur de thèse

Marion TELLIER, Professeure des universités, Université d'Aix-Marseille, Co-Directrice de thèse

Contrat de diffusion

Ce document est diffusé sous le contrat *Creative Commons* « Paternité – pas d'utilisation commerciale – pas de modification » : vous êtes libre de le reproduire, de le distribuer et de le communiquer au public à condition d'en mentionner le nom de l'auteur et de ne pas le modifier, le transformer, l'adapter ni l'utiliser à des fins commerciales.

Université Lumière Lyon 2

École doctorale Lettres, Langues, Linguistique et Arts - ED 484

ICAR (Interactions, Corpus, Apprentissages, Représentations) - UMR 5191

Séquences d'explication lexicale dans l'enseignement du français par visioconférence : une approche multimodale

Benjamin Holt

Thèse pour l'obtention du grade de Docteur en Sciences du Langage de l'Université Lumière Lyon 2

Sous la codirection de **Nicolas Guichon** et de **Marion Tellier**

Présentée et soutenue publiquement le 12 octobre 2018

Composition du jury :

Lucile Cadet, Professeure des universités, Université de Cergy-Pontoise — ÉSPÉ

Nicolas Guichon, Professeur des universités, Université Lumière Lyon 2

Heather Hilton, Professeure des universités, Université Lumière Lyon 2

Annick Rivens Mompean, Professeure des universités, Université Lille 3

Thierry Soubrié, Maître de conférences, Université Grenoble Alpes

Marion Tellier, Professeure des universités, Aix-Marseille Université

Remerciements

Cette thèse est le fruit d'un travail collaboratif et n'aurait pu voir le jour sans l'aide des personnes ci-après.

Je remercie tout d'abord mes directeurs de thèse, Nicolas Guichon et Marion Tellier, d'avoir accepté de travailler avec moi et pour leur encadrement tout au long de mon parcours doctoral. Ils ont toujours cru en moi, même en périodes de difficultés. Leurs relectures, propositions scientifiques, disponibilité, et encouragements m'ont permis de définir mon sujet et d'améliorer mes analyses. Leur accompagnement à des conférences en France et à l'étranger m'a permis de développer plusieurs compétences professionnelles et communicationnelles qui me serviront tout au long de ma carrière.

Je remercie les rapporteures, Lucile Cadet et Annick Rivens Mompean, pour leurs remarques judicieuses, ainsi que Heather Hilton et Thierry Soubrié pour leur participation en tant que membres du jury.

Cette thèse a été financée par le LABEX ASLAN (ANR-10-LABX-0081) de l'Université de Lyon dans le cadre du programme "Investissements d'Avenir" (ANR-11-IDEX-0007) de l'État Français géré par l'Agence Nationale de la Recherche (ANR). Merci à François Pellegrino et Kris Lund pour leur aide en tant que directeurs du Labex.

La partie audiovisuelle du corpus, c'est-à-dire les enregistrements vidéo sur lesquels portent les analyses, n'aurait pu être construite sans l'aide de l'équipe Cellule Corpus Complexes du laboratoire ICAR. Je tiens à remercier tout particulièrement Justine Lascar et Daniel Valéro. Justine Lascar m'a appris à utiliser plusieurs logiciels, notamment Final Cut Pro, ce qui m'a permis de construire des fichiers audiovisuels exploitables, et de construire les extraits qui font partie du site compagnon. Daniel Valéro m'a aidé à obtenir et à utiliser le matériel qui m'a permis de faire les analyses et de rédiger la thèse, et m'a aidé à construire le site compagnon. Je remercie également toutes les personnes qui ont commencé à construire le corpus avant mon arrivée à Lyon, notamment Nicolas Guichon, Françoise Blin, Cathy Cohen, Marie Josée Hamel, Kim Szczot, Sylvie Thouësny, Julie Vidal et Ciara Wigham. Je remercie toutes les personnes qui ont aidé à transcrire le corpus : Thomas Debay, Matthieu Decorde, Amina Dib, Lycia Dobé, Laurie Frosini, Angèle Meurgue Guyard, Shuo Jiang, Aline Pellicier, Jean-Guy Sebire, et Julie Vidal.

Les discussions stimulantes et l'accompagnement dont j'ai bénéficié lors des séminaires 3MA (Médias, Médiations, Multimodalité dans l'Apprentissage) m'ont aidé à approfondir ma thèse et je remercie tous les membres : Nicolas Guichon, Karine Bouchet, Cathy Cohen, Eugenie Duthoit, Jean-François Grassin, Erica Johnson, Salifou Koné, Elise Merlet, Joséphine Remon, Zoum Tou, Angelika Verch, Julie Vidal, et Ciara Wigham. Un grand merci à tous les invités qui sont intervenus au séminaire, en particulier Virginie Fasel-Lauzon qui est venue présenter ses travaux sur les explications lexicales en classe de FLE et dont le travail m'a grandement inspiré.

Je remercie le laboratoire ICAR (UMR 5191) de m'avoir accueilli et intégré. Merci à Sandra Teston Bonard pour sa direction, et à Agnès Bailly et Lucie Bujon pour l'aide administrative et logistique qu'elles m'ont apporté tout au long de mon parcours doctoral. Merci à la Commission des Doctorants, notamment à Isabel Colón de Carvajal, pour leur aide et suivi, et à Jean-Philippe Maitre pour sa relecture de l'un de mes chapitres méthodologiques.

J'ai eu la chance de pouvoir partager des articles, idées, extraits de corpus, analyses, et problèmes personnels avec les membres du « Journal Club » autrement dit le « Mickey Club » : Karine Bouchet, Aparajita Dey-Plissonneau, Morgane Domanchin, Salifou Koné, Elise Merlet, et Julie Vidal.

J'ai vécu des moments joyeux avec toutes les personnes avec qui j'ai eu le plaisir de partager un bureau au sein du laboratoire ICAR, et j'ai bénéficié de leur soutien, relectures et références bibliographiques : Morgane Domanchin, Gloria El Hajj, Safa Al Muhammad, Kevin Petit, Vanessa Piccoli, Mohammed Smail, Lucien Tisserand, Biagio Ursi, et Yiru Xu.

Je remercie Véronique Traverso de m'avoir permis de participer à l'école thématique MAINLY, ce qui m'a servi d'initiation à l'analyse conversationnelle.

Merci à Marco Cappellni et Christelle Combe Celik pour le feedback qu'ils m'ont donné lors de mon comité de suivi de thèse en fin de deuxième année au Laboratoire Parole et Langage (Aix-en-Provence). Je remercie également Cathy Cohen, Martine Marquillo et Véronique Rivière pour le suivi et feedback qu'elles m'ont apporté à Lyon en fin de troisième année.

Merci à François-Xavier Vautherin de l'association Coup de Pouce Université pour ses relectures et pour le soutien moral qu'il m'a apporté régulièrement pendant mes deux dernières années de thèse. Nos séances de travail ont profondément enrichi mes connaissances lexicales et culturelles et m'ont aidé à avancer dans la rédaction.

Je remercie Vivian Waltz et Christian Cavé de m'avoir guidé vers le FLE il y a sept ans. Sans leur aide, je ne serais probablement pas resté en France après 2011.

Je remercie mes parents, Tom et Bonnie, et mes grands-parents, Bob et Carmen Riggs, pour leur soutien tout au long de mes études en France. Merci à mon amie Emily Cania, pour son soutien tout au long de mon parcours doctoral.

Sommaire

Introduction 9
Partie I : Cadre théorique 17
Chapitre 1 : Le contexte 19
Chapitre 2 : Enseignement et apprentissage du vocabulaire en langue étrangère 37
Chapitre 3 : Vers une approche multimodale de l'analyse des séquences d'explication lexicale en visioconférence 75
Partie II : Méthodologie 113
Chapitre 4 : Construction du corpus 115
Chapitre 5 : Annotation des données et repérage des séquences d'explication lexicale 141
Partie III : Analyses des données 173
Chapitre 6 : La focalisation sur la forme des items lexicaux 175
Chapitre 7 : Aider l'apprenant à accéder au sens d'un item lexical 215
Chapitre 8 : Impliquer les apprenants 259
Chapitre 9 : Expliquer un même item lexical différemment 297
Chapitre 10 : Discussion 343
Conclusion 357
Bibliographie 365
Annexes 381
Glossaire 403
Table des matières 409

Introduction

Ancrage

Cette thèse est la continuation de notre mémoire de Master 2 que nous avons soutenu en juin 2014 à l'Université d'Aix-Marseille sous la direction de Marion Tellier. Intéressé par les études gestuelles, nous avons eu la chance, grâce à Nicolas Guichon et le laboratoire ICAR, de mener nos analyses sur une partie du corpus qui est devenu par la suite l'objet de cette thèse. Dans le cadre de notre mémoire de Master 2, nous avions analysé la production gestuelle de quatre futurs enseignants de français langue étrangère en visioconférence afin de voir si les gestes étaient utilisés davantage lors des séquences d'incompréhension que lors des séquences dites normales[1]. Nos résultats avaient montré que ce n'était pas le cas. Nous avons mentionné, en conclusion de cette étude, qu'il serait intéressant d'étudier les *fonctions* des gestes produits par les enseignants en visioconférence et non simplement leurs *catégories* (voir annexe 1 en page 381), et de mener des études sur plusieurs semaines d'interaction. C'est alors que l'opportunité de continuer à travailler sur le même corpus à l'Université Lumière Lyon 2 s'est présentée[2].

Avec un corpus aussi large, de nombreuses entrées didactiques étaient possibles. Après une année de réflexion pendant laquelle nous avons contribué à la construction du corpus, nous avons décidé d'entrer dans les données par le lexique pour des raisons personnelles et didactiques.

Intérêt personnel pour l'apprentissage du vocabulaire

L'acquisition du vocabulaire n'est jamais terminée. Cependant, l'acquisition est plus rapide pendant certaines périodes de la vie que d'autres. En effet, les enfants à l'école primaire apprennent en moyenne mille mots par an dans leur langue maternelle (Biemiller & Slonim, 2001). En langue étrangère, les mots de haute fréquence sont appris plus rapidement que les mots plus rares (Nation, 2013).

[1] Nous avons défini une séquence « normale » comme une séquence dans laquelle aucune incompréhension n'a été détectée.

[2] Thèse financée par le LabEx ASLAN de l'Université de Lyon.

[3] Club Sportif Pertuisien

Ma connaissance des mots les plus fréquents en français était toujours en voie de développement quand j'ai passé trois mois en Normandie en 2008 en tant qu'étudiant en licence. À vingt ans, c'était mon premier séjour en France. J'étais hébergé dans une famille d'accueil particulièrement bienveillante quand certains moments d'apprentissage du lexique ont prêté à rire. Alors que ma « mère d'accueil » me faisait visiter un étang dans lequel il y avait des *canards*, j'ai dit « regardez les *connards* ». Elle m'a corrigé en riant. Plus tard lors d'un dîner, alors que l'on mangeait des compotes pour le dessert, j'ai dit « passez-moi les capotes ». Tous les membres de la famille ont éclaté de rire. Grâce à ces expériences, j'ai pu découvrir le côté ludique et parfois comique de l'apprentissage du vocabulaire.

C'est grâce à cette « mère d'accueil » que j'ai développé le goût pour la collaboration intense, étroite et multimodale qui se met en place entre apprenant et enseignant lors des séquences d'explication lexicale. En faisant des gestes pour remplir des lacunes lexicales, j'ai pu obtenir la forme correcte de plusieurs mots : le mot *flèche* quand nous circulions sur la route et j'ai souhaité connaître le mot pour parler des flèches de rabattement. J'ai ressenti des sentiments de satisfaction après chaque séquence d'explication lexicale réussie. J'ai aussi compris l'importance du rappel (Baddeley, 1993 ; Laufer & Hulstijn, 2001 ; Nation & Webb, 2011) quand ma « mère d'accueil » m'a demandé, une semaine après l'apprentissage, comment dire *girouette* en pointant du doigt une girouette sur une église. Elle m'a permis d'apprécier l'effet stimulant que peut avoir le partage de connaissances quand elle m'a demandé la traduction du mot *essuie-glace* en anglais après m'avoir appris la forme française. C'est lors de ce même séjour que j'ai vécu un moment particulièrement mémorable. En effet, ne connaissant pas le mot en français pour *snowmobile*, je l'ai deviné, en faisant ce que Lüdi (1994) appelle le *bricolage lexical*, pour dire « motoneige ». J'étais surpris et content quand ma « mère d'accueil » m'a dit que c'était la forme correcte.

Après avoir appris la plupart des mots de haute fréquence en français, mes connaissances lexicales ont connu une période d'apprentissage particulièrement accéléré lors de mes deux années de Master (de 2012 à 2014). Grâce à mon contact avec des milieux différents, j'ai pris du plaisir à apprendre plusieurs mots et expressions appartenant à des registres différents. Pendant la journée, grâce aux cours à l'Université d'Aix-Marseille, j'étais en contact avec le langage universitaire. En même temps, j'étais adhérent à une association de sport[3] à Pertuis[4] où j'habitais. J'étais là en contact avec un tout autre registre du français que

[3] Club Sportif Pertuisien

[4] Pertuis est une ville de 20 000 habitants située aux alentours d'Aix-en-Provence.

l'on pourrait qualifier de « populaire ». Une de mes professeurs de Master employait de temps en temps l'expression *je m'assois dessus*. L'expression équivalente que j'ai apprise dans la salle de sport appartient à un autre registre du français et ne serait pas la bienvenue en milieu universitaire.

Le président du Club Sportif Pertuisien, un homme âgé de plus de 80 ans avec qui je suis devenu ami, m'a particulièrement marqué et inspiré, aussi bien par sa forme physique que par ses connaissances lexicales. En effet, il maîtrisait *tous* les registres du français, du vulgaire au plus soutenu, et se réjouissait de me faire part de ses vastes connaissances. Nos conversations régulières nous ont permis de bénéficier de nombreuses séquences d'explication lexicale. C'est grâce à lui que j'ai appris plusieurs mots en latin, par exemple *vulgum pecus*, et quelques mots en provençal. C'est aussi grâce à lui que je connais une dizaine d'expressions équivalentes à *je m'assois dessus*, y compris ma préférée, *peu me chaut*. Suite à ces expériences personnelles, sans mentionner tout ce que j'ai appris au cours de cette thèse, et sans mentionner toute l'acquisition lexicale dont j'ai bénéficié grâce à la lecture, je suis d'avis que l'acquisition incidente du vocabulaire par interaction est une activité stimulante, souvent ludique, parfois comique, et toujours enrichissante.[5]

Intérêts didactiques et pratiques pour l'entrée par le lexique

L'apprentissage du lexique est fondamental pour tout apprenant d'une langue étrangère. Alors que la communication verbale orale est possible sans connaissances grammaticales, elle ne l'est pas sans connaissances lexicales (Folse, 2004). Bien que l'acquisition lexicale soit « la base même de l'apprentissage linguistique », les méthodes d'enseignement restent sous-développées (Hilton, 2002 : 1)[6]. Cortier & Faugeras (2004 : 75) soutiennent que « le lexique est le parent pauvre de l'enseignement de la langue » et que les compétences lexicales chez les apprenants de FLE sont souvent insuffisantes. Nous expliquerons dans le cadre théorique de notre thèse pourquoi l'enseignement explicite du vocabulaire est indispensable, surtout pour les niveaux débutant et intermédiaire.

Une entrée par le lexique permet d'aborder une grande variété de phénomènes. Le lexique est étroitement lié non seulement à la grammaire et à la syntaxe mais aussi à la culture. L'entrée par le lexique permet donc d'analyser la façon dont plusieurs aspects de la

[5] Nous reprenons désormais le code académique. Le « nous » sera donc adopté. Le « nous » académique fait référence aux chercheurs et à la communauté des chercheurs du projet ISMAEL.

[6] Cette pagination correspond au manuscrit auteur.

langue sont expliqués par les enseignants. Une analyse d'explications portant sur des mots à forte charge culturelle (Galisson, 1987, 1991) sera proposée.

Comme évoqué ci-dessus, notre mémoire de Master 2 portait sur les gestes pédagogiques (Tellier, 2008a) produits par les futurs professeurs de langue étrangère en formation. Nous avions comptabilisé la production de chaque type[7] de geste lors des séquences d'incompréhension et lors des séquences dans laquelle aucune incompréhension n'était observée. Notre choix de nous concentrer en particulier sur les gestes venait du fait que le champ de la webcam limite profondément l'espace gestuel[8], et que par conséquent, selon nous, les gestes qui étaient rendus visibles étaient conscientisés à cause de l'effort requis pour les produire dans le champ de la webcam. Contrairement à nos hypothèses, nos résultats avaient montré que les gestes n'étaient pas produits davantage lors des séquences d'incompréhension que lors des séquences normales. Il s'était avéré de surcroît, que la plupart des gestes avaient été produits hors champ et étaient donc invisibles pour les participants (Guichon & Wigham, 2016). Pour ces raisons, une analyse qui porte uniquement sur la production gestuelle des participants ne semble pas être un bon point de départ pour une thèse.

Nous nous sommes donc tourné vers la multimodalité, qui prend en compte *toutes* les ressources sémiotiques, y compris les gestes. Une approche multimodale adopte une vue holistique de la communication où la parole, les images, les expressions faciales, les regards, les postures, et les gestes travaillent tous ensemble pour communiquer[9]. A l'instar de Nicolaev (2012) qui s'est intéressée aux séquences métalangagières, nous nous sommes posé la question suivante : comment les ressources sémiotiques sont-elles orchestrées par les enseignants lors des séquences de négociation du sens[10] ? C'est en consultant la littérature sur la négociation du sens que nous nous sommes rendu compte que la plupart des séquences de négociation du sens sont déclenchées par des items lexicaux. Cela veut dire que les items lexicaux sont les déclencheurs par excellence des séquences potentiellement

[7] Nous avons adopté la classification de gestes de McNeill (1992) à quatre dimensions : les iconiques, les métaphoriques, les déictiques, et les battements. A ces catégories nous avons ajouté les iconiques (Ekman et Friesen, 1969). Voir annexe 1 en page 381.

[8] Une réflexion plus approfondie sur l'espace gestuel sera présentée dans le chapitre 5.

[9] Notre conception de la multimodalité sera décrite dans le chapitre 3.

[10] Les séquences de négociation du sens (cf. chapitres 2 et 3) sont des moments où un élément langagier devient objet de discours. Ces moments sont propices à l'acquisition langagière.

acquisitionnelles[11] (de Pietro, Matthey, & Py, 1989). L'entrée par le lexique nous semble donc appropriée pour l'étude de la multimodalité et des SPA.

Objectifs

Nos objectifs sont à la fois descriptifs et praxéologiques. Nous viserons donc dans un premier temps à **comprendre la façon dont les items lexicaux sont expliqués multimodalement par les enseignants en visioconférence.** Quelles ressources sont mobilisées pour expliquer les facettes[12] de la connaissance de l'item lexical ? Nous mènerons ces analyses afin de formuler quelques propositions didactiques à destination de praticiens et de formateurs de futurs enseignants de langues étrangères en visioconférence.

Un autre objectif de notre thèse est lié à **la construction et à l'annotation d'un corpus audiovisuel.** La construction d'un corpus audiovisuel nécessite plusieurs transformations des données à plusieurs niveaux. Une réflexion sur ces transformations sera proposée dans la partie méthodologique de cette thèse. La construction du corpus dépasse le cadre de cette thèse et a nécessité la collaboration d'une grande équipe de chercheurs. Ce travail d'équipe a donné lieu notamment à la publication d'un ouvrage (Guichon & Tellier, 2017).

Enfin, concernant l'annotation des données, nous pensons que notre façon d'annoter des phénomènes multimodaux, notamment les gestes, permettra à d'autres chercheurs de s'en s'inspirer. L'annotation des gestes en visioconférence n'est pas semblable à l'annotation des gestes en présentiel pour diverses raisons qui seront abordées dans la partie méthodologique de cette thèse (voir chapitre 5).

Méthode

Nous saisirons des aspects de l'acte d'enseignement à partir d'interactions didactiques et écologiques (Cicurel, 2011). Nous adopterons une démarche inductive dans le sens où nous ne faisons pas d'hypothèses de départ (Ten Have, 1999). Cependant, notre démarche n'est pas strictement inductive, car nous avons une définition générale du phénomène recherché avant de procéder au repérage des séquences d'explication lexicale dans le corpus. Notre objectif

[11] Les séquences potentiellement acquisitionnelles sont des moments dans l'interaction entre natifs et non natifs où on observe une concentration de processus centraux de l'acquisition langagière (de Pietro, Matthey, & Py, 1989).

[12] De manière très générale, la connaissance d'un item lexical peut être divisée en trois facettes : la forme, le sens, et l'usage. La connaissance d'un item lexical sera abordée dans le chapitre 2.

n'est donc pas de trouver une définition d'une explication lexicale, mais de comprendre de quelle façon les ressources sémiotiques sont déployées par les enseignants lors des séquences d'explication lexicale et pour quelles fonctions pédagogiques.

Nous adopterons une approche multimodale selon laquelle aucune ressource sémiotique ne prime *a priori* sur une autre, et toutes les ressources contribuent à produire du sens (Jewitt, 2014). Nous adopterons cette approche par contraste aux méthodes plus classiques inspirées par l'analyse de la conversation qui donnent la priorité au langage verbal et qui donnent une place secondaire aux autres ressources aussi importantes comme les gestes, les expressions faciales, et les postures. Une approche multimodale est d'autant plus intéressante, voire nécessaire, pour analyser des interactions en visioconférence que les interactants disposent d'au moins trois canaux de communication : un microphone, une webcam, et un clavier.

Plan

Cette thèse est composée de trois parties : un état de l'art, une partie méthodologique, et une partie d'analyses.

La première partie se veut théorique et est composée de trois chapitres. Dans le premier chapitre, nous décrirons le contexte dans lequel s'inscrit notre projet de recherche. Le contexte sera présenté en premier lieu afin de donner au lecteur une vision globale de la lignée de recherche dans laquelle s'inscrit ce projet, de la plateforme de visioconférence utilisée, et des participants. Dans le deuxième chapitre, nous parlerons de l'apprentissage et de l'enseignement du lexique dans une langue étrangère afin d'en comprendre les enjeux. Dans le troisième chapitre, nous argumenterons en faveur d'une approche multimodale pour analyser les séquences d'explication lexicale en visioconférence.

La partie méthodologique est composée de deux chapitres. Dans le premier chapitre (4), nous expliquerons la façon dont le corpus a été construit en collaboration avec plusieurs personnes. Le rôle que nous y avons joué sera mis en valeur. Dans le deuxième chapitre (5), nous élaborerons notre positionnement épistémologique par rapport à nos données avant de décrire l'annotation de celles-ci et le repérage des séquences d'explication lexicale qui feront l'objet des analyses dans la troisième partie.

La troisième partie est composée de quatre chapitres d'analyses, et un chapitre synthétique de discussion. Les deux premiers chapitres (6 et 7) analyseront les moyens par lesquels les apprentis-enseignants se focalisent sur la forme et sur le sens des items lexicaux.

Les fonctions pédagogiques des différents types de focalisation sur la forme et sur le sens seront mises en valeur. Le troisième chapitre (8) s'intéressera à la façon dont les apprentis-enseignants assurent l'implication cognitive et socioaffective des apprenants lors des séquences d'explication lexicale. Le quatrième chapitre (9) profitera d'une opportunité assez rare. Nous pourrons comparer cinq apprentis-enseignants dans des conditions similaires expliquant les mêmes items lexicaux portant une charge culturelle importante à leurs apprenants. Dans le dernier chapitre (10), nous mènerons une réflexion sur nos questions de recherche, nos analyses, nos résultats, et l'apport de cette thèse à l'analyse multimodale et à la formation des enseignants.

Partie I : Cadre théorique

Dans cette partie, nous ferons un état de l'art sur l'enseignement des langues étrangères en visioconférence, l'enseignement du vocabulaire en langue étrangère, et la multimodalité afin de nous doter d'outils théoriques nécessaires à l'analyse multimodale de séquences d'explication lexicale en visioconférence.

Le premier chapitre présentera un bilan des avancées technologiques qui ont permis l'émergence de l'enseignement des langues à distance et de leur étude. Il visera ensuite à situer notre projet de thèse par rapport à d'autres projets de recherche qui s'intéressent à l'enseignement des langues étrangères à distance. Nos données seront présentées de manière synthétique sans rentrer dans les détails de leur construction afin d'esquisser des questions de recherche.

Le deuxième chapitre justifiera la nécessité d'un enseignement explicite du vocabulaire en langue étrangère. Puis, en nous appuyant sur des études expérimentales, nous ferons état des procédés par lesquels les enseignants peuvent favoriser la transmission de connaissances lexicales. Nos questions de recherche préliminaires seront posées.

Le troisième chapitre présentera une modélisation classique des séquences de négociation du sens et des séquences d'explication lexicale. Cette présentation mettra en évidence l'absence d'une prise en compte de la dimension multimodale de la parole. Puis, nous argumenterons en faveur d'une approche multimodale de l'analyse des séquences d'explication lexicale en visioconférence tout en définissant quelques notions clés de la multimodalité. Nos questions de recherche seront approfondies.

Chapitre 1 : Le contexte

1.1 Introduction

Cette thèse s'inscrit dans un courant de recherche qui a émergé avec la disponibilité d'Internet à grande échelle. L'objectif de ce chapitre est de positionner notre thèse par rapport à d'autres projets de recherche dans ce même domaine. Les études qui ont été menées sont trop nombreuses pour permettre une revue exhaustive de la littérature sur ce thème. Nous allons citer uniquement les études qui nous semblent les plus pertinentes pour notre recherche et qui sont les plus à même d'aider à comprendre le contexte dans lequel se situe cette thèse.

Nous résumerons dans un premier temps les innovations techniques qui ont permis l'émergence de dispositifs d'enseignement à distance comme la visioconférence. Ce faisant, nous décrirons les changements pratiques et pédagogiques induits par l'enseignement des langues étrangères en visioconférence. Puis, nous présenterons le projet de recherche dans laquelle cette thèse s'inscrit. L'utilité scientifique du corpus sera discutée, ce qui nous donnera l'occasion d'aborder des questions de recherche. Nous pouvons d'ores et déjà présenter des dates importantes de la lignée de recherche dans laquelle nous nous inscrivons.

Dates	Évènements
Années 1990	Développement de supports pédagogiques en format CD-ROM
Années 1990	Développement du forum et du clavardage
2000	Publication de l'article "Intercultural learning via videoconferencing" (Apprentissage interculturel par visioconférence) par O'Dowd
2002	Création du Français en Première Ligne (F1L)
2003	Sortie de Skype
2006	Début du F1L en mode synchrone
2008	Publication de l'article « Allo Berkeley ? » par Develotte, Guichon et Kern
2010	Développement de Visu
2011	Publication de l'ouvrage « Décrire la conversation en ligne » édité par Develotte, Kern et Lamy
2011	Soutenance de thèse de doctorat de Samira Drissi
2012	Soutenance de thèse de doctorat de Caroline Vincent
2012	Soutenance de thèse de doctorat de Viorica Nicolaev
2013	Enregistrement d'un semestre d'échanges entre Lyon et Dublin
2014	Soutenance de thèse de doctorat de Tatiana Codreanu
2017	Publication de l'ouvrage « Enseigner l'oral en ligne » édité par Guichon et Tellier

Figure 1 : Dates clés dans l'enseignement du français en visioconférence

1.2 L'enseignement et l'apprentissage des langues médiés par ordinateur

Grâce à leur potentiel comme supports pédagogiques, les ordinateurs sont utilisés comme des outils pédagogiques dans diverses disciplines académiques, y compris celle de l'enseignement et l'apprentissage des langues étrangères, depuis plus de trente ans. Les ordinateurs sont susceptibles d'augmenter les potentiels d'apprentissage par rapport aux supports classiques comme les livres. Mangenot & Moulin (2000 : 122) soutiennent que les technologies de l'information et de la communication (désormais TIC) peuvent aider les apprenants à acquérir du vocabulaire car « celles-ci, grâce notamment aux fonctions de recherche et à l'hypertexte, permettraient de réconcilier langue et discours, de contextualiser les unités lexicales dans des corpus comportant des textes authentiques d'origines variées ». L'hypertexte permet aux apprenants de consulter des définitions de mots inconnus plus rapidement que par un dictionnaire classique, ce qui permet à ceux-ci de consulter des mots qui seraient autrement ignorés.

Si les supports pédagogiques en format CD-ROM pendant les années 90 ont offert à des enseignants et à des apprenants un support technique qui était une évolution par rapport aux supports classiques comme les livres[1], c'est Internet qui a suscité de véritables changements sur nos façons d'apprendre et d'interagir. La disponibilité d'Internet a donné naissance à plusieurs types de *communications médiées par ordinateur* (CMO).

1.2.1 L'évolution des TIC

Dans cette section nous allons résumer les avancées technologiques qui ont abouti à l'émergence de la visioconférence telle que nous la connaissons aujourd'hui. Le tableau ci-dessous résume les affordances[2] offertes par chaque type d'outil. Une affordance relie l'acteur et son environnement (Gibson, 1979). Les affordances de l'environnement englobent tout ce qui est offert à l'acteur, autrement dit ce qu'un environnement permet à un acteur de faire. Blin (2016) rappelle que les affordances existent qu'elles soient perçues ou non. Par exemple, la visioconférence permet à ses utilisateurs de parler et de clavarder en même temps, tout en voyant l'image de l'interlocuteur.

[1] Les supports didactiques sur ordinateur offrent de nombreux avantages. En lisant un texte sur un écran, on peut, par exemple, accéder directement à une définition écrite ou à une image en mettant le curseur sur un mot non compris (Plass, Chun, Mayer, & Leutner, 1998).

[2] Affordance est un terme anglais qui signifie capacité, possibilité.

	Outils	Caractéristiques techniques
Asynchrone	Forum	Les forums permettent aux participants de créer des fils de discussion et de poster des messages. L'historique du fil de discussion est visible à tout moment. Les participants peuvent se connecter et poster quand ils le veulent.
Synchrone	Clavardage	Les dispositifs de clavardage permettent aux participants d'envoyer des messages qui apparaissent instantanément sur l'écran de l'autre. Les participants doivent être connectés tous en même temps. L'historique de l'échange est visible à tout moment.
	Audioconférence	Avec un microphone et un clavier, les participants peuvent communiquer oralement, comme au téléphone, et par clavardage simultanément. L'historique de l'échange textuel reste visible, mais l'audio est éphémère.
	Visioconférence	Les plateformes de visioconférence offrent les mêmes affordances que les plateformes d'audioconférence mais permettent aux participants, grâce à une webcam, de se voir. Certaines plateformes permettent aux participants d'envoyer des documents audiovisuels comme des photos et des vidéos.

Figure 2 : Évolution des TIC

Quand l'Internet était relativement nouveau, le débit de connexion n'était pas assez élevé pour permettre des échanges visiophoniques. La communication textuelle *asynchrone* était alors privilégiée. Lors de la communication asynchrone, il existe un décalage important entre l'émission et la réception du message. Le forum en est un bon exemple. Les forums permettent aux apprenants de publier et de lire à tout moment sur des thèmes divers et d'appréhender tout l'historique de ce qui a été dit (Mangenot, 2008). L'avantage de la communication asynchrone permet aux apprenants de se connecter quand ils le veulent, et leur donne le temps de bien réfléchir avant de produire du langage (Brudermann, 2010 ; Mangenot, 2008 ; Stockwell, 2010). La communication asynchrone est utilisée non seulement pour les cours à distance mais aussi pour des dispositifs d'enseignement hybrides (Charlier, Deschryver, & Peraya, 2006 ; Degache & Nissen, 2008 ; Nissen, 2006 ; Soubrié, 2008) qui se caractérisent par une articulation entre enseignement en présentiel et enseignement à distance.

La communication *synchrone* a lieu quand un message est reçu instantanément après avoir été envoyé. La communication textuelle par clavardage s'est répandue pendant les années 90 et a fait l'objet de nombreuses études en didactique des langues. Le clavardage semble favoriser la participation égale des apprenants (Warschauer, 1996) ainsi que développement de l'interlangue et de compétences interactionnelles en langue étrangère (Tudini, 2003). L'émergence d'un sentiment de présence à distance est aussi privilégiée par le

clavardage, même si les participants ne voient jamais l'image de l'autre (Walther, Slovacek, & Tidwell, 2001). Enfin, l'interaction pédagogique par clavardage donne lieu à de nombreuses séquences de négociation du sens, ce qui favorise l'acquisition du vocabulaire (Smith, 2003).

La communication synchrone par audioconférence, et surtout par visioconférence, nécessite un accès à l'Internet à haut débit. Pour cette raison, l'utilisation de la visioconférence par des institutions scolaires est plus récente que l'utilisation de la communication asynchrone. L'interaction en visioconférence était possible avant les années 2000 (Goodfellow, Jefferys, Milest, & Shirra, 1996), mais ce n'était pas une technologie de tous les jours. Il n'était pas envisageable d'avoir une webcam et un écran par participant quand O'Dowd (2000) a mené une étude auprès d'apprenants d'espagnol dans une université américaine communiquant avec des apprenants d'anglais en Espagne. A partir du constat que les technologies permettent des échanges culturels, ce chercheur a mis en place un projet où chaque groupe a créé une vidéo décrivant sa ville. Puis, les groupes en ont discuté en visioconférence, groupe à groupe, pendant trois séances d'une heure. Cette étude a permis à ce chercheur de constater un gain de confiance chez les apprenants et de créer une liste de bonnes pratiques en visioconférence.

La visioconférence est devenue accessible au grand public pendant les années 2000 grâce aux logiciels comme Skype en 2003. Dès lors, tout utilisateur possédant un ordinateur, un microphone, une webcam, et une connexion Internet assez rapide pouvait communiquer par visioconférence. L'émergence de cette technologie a suscité des changements sur les façons dont les personnes communiquent. Plusieurs types de communications qui se faisaient auparavant par téléphone ou par d'autres moyens de communication ont commencé à évoluer grâce à l'utilisation de cet outil. Ces changements ont suscité rapidement l'intérêt de chercheurs travaillant dans différents domaines. Colón de Carvajal (2006), dans une démarche de l'analyse de la conversation, a étudié la co-présence et l'évolution du cadre participatif entre des patients sortant de l'hôpital et des employés d'un centre d'écoute lors d'interactions visiophoniques. Develotte et al., (2011) dans l'esprit d'un ouvrage sorti vingt ans plus tôt (Kerbrat-Orecchioni & Cosnier, 1987), se sont attachés à décrire les effets de la visioconférence sur la conversation ordinaire. En didactique des langues, la visioconférence est utilisée depuis 2006 pour enseigner le français langue étrangère.

1.2.2 Spécificités de l'enseignement par visioconférence

Par rapport à l'enseignement « classique » en présentiel, la visioconférence apporte plusieurs possibilités. Le temps de prise de parole chez les apprenants est beaucoup plus important que dans une salle de classe classique grâce au fait que ces derniers sont souvent seuls ou en binôme face à un enseignant. Puisque l'écart d'âge est faible dans les cas étudiés, les conversations ont un ton plus informel et il y a plus de connivence (Develotte & Drissi, 2013; Guichon & Drissi, 2008). La visioconférence permet aux apprenants de développer des compétences ethnographiques et de découvrir des cultures étrangères (O'Dowd, 2005). La visioconférence se prête particulièrement bien à des séquences de focalisation sur la langue (Sarré, 2011). Cependant, enseigner en présentiel dans une salle de classe n'est pas tout à fait pareil à enseigner en visioconférence.

L'enseignement en ligne provoque des ruptures par rapport à l'enseignement classique en salle de classe : ruptures spatiales, médiatiques, et temporelles (Develotte & Mangenot, 2007). Ces ruptures ont un impact sur la communication, et donc sur la gestion pédagogique nécessaire. Il faut non seulement gérer le côté didactique, mais aussi les côtés technique et socioaffectif. Le tableau ci-dessous résume certaines compétences requises des enseignants à cause des spécificités de la visioconférence.

Caractéristiques techniques	Compétences requises
Multicanalité	Maîtriser la polyfocalisation de l'attention
Images floues et mouvements saccadés	Bien se positionner devant la webcam, rendre ses gestes visibles et ne pas faire de mouvements brusques
	Donner des indices socioaffectifs et veiller au bien-être émotionnel des apprenants
Coupures intempestives	Savoir passer d'un canal à un autre à la suite d'une panne technique et maintenir la communication
Décalage entre émission et réception du message	Savoir gérer les tours de paroles
Distance physique entre la webcam et l'écran	Prendre conscience du fait que quand on regarde l'image de l'autre, on lui donne l'impression de ne pas être regardé
Temps de séance limité	Gérer les tâches, savoir diviser le feedback entre corrections immédiates et bilans écrits

Figure 3 : Compétences requises pour enseigner en visioconférence

Les plateformes de visioconférence sont *multicanales*[3], c'est-à-dire qu'elles disposent d'un microphone pour la parole orale, d'un système de clavardage pour le langage écrit, et d'une webcam pour la communication visuelle (gestes, mimiques faciales, hochements de tête, etc.). Les utilisateurs doivent apprendre à gérer ces outils et développer ce que Lamy (2006) nomme une *alphabétisation médiatique*. La gestion de plusieurs outils à la fois nécessite une certaine *polyfocalisation* de l'attention (Develotte & Drissi, 2013 ; Jones, 2004). A cette polyfocalisation sur ces outils techniques s'ajoute la dimension langagière de l'échange. A l'instar de Bange (1992), qui parle de *bifocalisation* sur l'activité et sur la langue, nous proposons la notion de *trifocalisation* : focalisation sur l'activité en cours, focalisation sur la langue, et focalisation sur la manipulation de l'outil technique.[4]

Bien qu'offrant des possibilités riches de communication par rapport à d'autres moyens de communication à distance plus rudimentaires, la visioconférence est un outil de médiation qui ne restitue pas parfaitement le son et l'image de l'autre : images floues, mouvements saccadés, désynchronisation entre parole et image, coupures intempestives constituent autant de phénomènes auxquels doivent faire face les utilisateurs de l'outil. Une maîtrise de la technologie devient donc nécessaire, faute de quoi la focalisation sur les opérations techniques primera sur la focalisation sur l'interaction pédagogique (Guichon & Drissi, 2008).

Les spécificités techniques du dispositif ont des répercussions sur la communication. Les interactants doivent se positionner devant la webcam et le microphone de manière à être vus et entendus. A la suite d'une panne technique, l'enseignant doit savoir assurer la continuité de la séance (Azaoui, 2017). Les interactants doivent apprendre à gérer les tours de parole (Traverso, 2011), ce qui n'est pas évident quand il existe un léger décalage entre l'émission et la réception du son, et quand les regards sont moins perceptibles qu'en présentiel. Concernant les regards, le contact oculaire direct n'est pas possible car pour donner l'impression à son interlocuteur qu'il est vu, il faut regarder directement vers la webcam au lieu de regarder l'image de l'autre (De Chanay, 2011).

Enseigner en visioconférence nécessite aussi la maîtrise de certaines compétences pédagogiques : les consignes, les explications, les rétroactions, les alertes et les réparations (Guichon & Drissi, 2008) et les tâches (Guichon & Nicolaev, 2009). En outre, il faut savoir

[3] La multicanalité permet la juxtaposition de parole, images, et texte (Lancien, 1998). La multicanalité et la multimodalité seront discutées ultérieurement dans le chapitre 3.

[4] Fasel Lauzon (2014) utilise le terme « trifocalisation » pour parler de la focalisation sur le contenu, sur la langue, et sur la tâche en cours.

gérer les passages entre interactant et correcteur, et savoir quel mode (cf. chapitre 3) utiliser et à quel moment (Vidal & Wigham, 2017). La gestion des tâches et du feedback est d'autant plus importante que le temps de la séance est limité à quarante minutes. Ce partage d'attention entre le côté technique de l'échange et le côté humain et pédagogique nécessite que les interactants en visioconférence apprennent à gérer plusieurs choses à la fois.

Enseigner en ligne nécessite que l'on se soucie du sentiment de présence en ligne que l'on crée (Jézégou, 2010). Puisque la webcam ne montre qu'une image dégradée de l'autre, il est important de donner des indices supplémentaires d'ordre culturel et socioaffectif (Guichon & Wigham, 2016). Les enseignants en ligne doivent veiller au bien-être émotionnel des apprenants. Jézégou (2010) nomme trois types de présence à distance : présence cognitive, socio-affective, et pédagogique. La présence pédagogique est nécessaire pour la présence socio-affective, qui est à son tour nécessaire pour la présence cognitive. Satar (2015), à partir d'analyses qualitatives, définit des stratégies pour soutenir l'interaction, notamment poser des questions, trouver des sujets intéressants, sourire, fournir des backchannels et des compliments, et simplifier le langage. Guichon (2017a) souligne l'importance de l'écoute à l'autre, la connivence et la symétrie lors des interactions en visioconférence.

La visioconférence est donc particulière en raison de la multitude de canaux offerts (webcam, microphone, clavier) et pour la polyfocalisation qui est requise en conséquence. Les enseignants de langue en ligne doivent en conséquence développer des *compétences sémio-pédagogiques* (Guichon, 2013) afin de gérer les côtés technique, pédagogique et humain des échanges. Une autre particularité de visioconférence qui aide au développement de ces compétences est la possibilité d'enregistrer les interactions de manière peu intrusive via un logiciel de capture d'écran qui enregistre tout ce qui se passe sur l'écran. La rétrospection facilite la réflexivité des apprenants, ce qui est une pratique fondamentale au développement professionnel des enseignants (Soubrié, 2016). Les captures d'écran seront discutées dans le chapitre 4. Puisque l'enseignement en ligne a un statut particulier, de nouveaux courants de recherche se sont développés pour essayer de comprendre et d'améliorer ce type d'enseignement, y compris celui dans lequel s'inscrit notre thèse.

1.2.3 Le Français en Première Ligne

Le Français en Première Ligne (F1L), créé par Christine Develotte et François Mangenot, existe depuis 2002 et est synchrone depuis 2006 (Codreanu & Celik, 2013; Codreanu & Develotte, 2010 ; Develotte & Drissi, 2013). Ce dispositif permet à des

apprenants de français langue étrangère d'interagir à distance avec des francophones natifs, et permet à de futurs enseignants de français langue étrangère (désormais FLE) de s'entraîner à l'enseignement et à l'usage de la technologie. Le projet a débuté en mode asynchrone avec l'usage de forums. Dans le cadre de la version synchrone de ce dispositif, développé en collaboration avec Nicolas Guichon, de futurs professeurs inscrits en Master FLE à Lyon ont donné des cours de français à des apprenants de niveau intermédiaire à l'Université de Californie à Berkeley en 2007. Develotte, Guichon, & Kern, (2008 : 150) ont trouvé alors que la visioconférence avait du potentiel « non seulement pour l'apprentissage des langues étrangères mais aussi pour la formation des professeurs de langues ». C'est dans cette optique de recherche que notre thèse s'inscrit.

Le F1L a donné lieu à quantité de projets de recherche, et le nôtre s'inscrit dans cette lignée de recherche. De manière non exhaustive, il a été question d'étudier les différents types de régulations pédagogiques en visioconférence (Guichon & Drissi, 2008), les compétences requises pour enseigner une langue étrangère en visioconférence (Dejean-Thircuir, Guichon, & Nicolaev, 2010 ; Guichon, 2009), les effets de différents types de tâches sur la production orale des apprenants (Guichon & Nicolaev, 2009), l'utilisation de la webcam et des autres canaux du dispositif (Develotte, Guichon, & Vincent, 2010 ; Guichon & Wigham, 2016), la façon dont les enseignants en formation donnent du feedback (Guichon et al., 2012), le regard critique que développent les enseignants en formation sur leur propre pratique d'enseignement (Guichon, 2013), la création de bilans suite aux séances d'interaction (Rivière & Guichon, 2014), la production de gestes lors de séquences d'incompréhension (Holt, Tellier, & Guichon, 2015), et l'explication d'items lexicaux (Wigham, 2017).

Une vision globale de ces études se trouve en annexe 2 en page 383. Ces études montrent que l'interaction pédagogique en visioconférence rend possible le traitement de plusieurs questions de recherche variées. Nous constatons que les questions de recherche deviennent progressivement plus pointues. Alors que l'étude de Develotte, Guichon, & Kern (2008) a été exploratoire avec des résultats très généraux, les études suivantes répondent à des questions de recherche beaucoup plus précises. Notre thèse, ainsi que celles de Julie Vidal et d'Aparajita Dey-Plissonneau, sont dans le prolongement de cette lignée de recherche (voir annexe 2 en page 383).

1.3 Le projet ISMAEL

Notre thèse s'inscrit dans le cadre du projet ISMAEL (InteractionS et Multimodalité dans l'Apprentissage et Enseignement d'une Langue), projet cogéré par l'équipe ICAR[5] à Lyon, France et l'équipe Dublin City University (DCU) en Irlande (Guichon & Tellier, 2017). Ce projet a été codirigé par les formateurs de chaque université : Françoise Blin à DCU et Nicolas Guichon à l'Université Lumière Lyon 2. Les objectifs du projet ISMAEL sont doubles : il s'agit d'étudier le jeu de la multimodalité dans les interactions pédagogiques en ligne et de construire et partager un corpus avec plusieurs chercheurs de différentes universités possédant des compétences et des questions de recherche variées. Le premier objectif s'aligne avec les questions de recherche de ce travail de thèse, et le second nécessite un corpus qui soit utilisable, consultable et partageable (Cohen & Guichon, 2016). La construction du corpus sera détaillée dans le chapitre 4. Nous décrirons ici les dispositifs pédagogiques et techniques afin d'aborder les questions didactiques dont le traitement est rendu possible par le projet et dont nos propres questions de recherche émergent.

1.3.1 Le dispositif pédagogique

Lors du premier semestre de l'année universitaire 2013-2014, des étudiants en deuxième année de Master FLE professionnel de Nicolas Guichon à Lyon enseignaient le français langue étrangère à des apprenants de Françoise Blin à Dublin. Les étudiants à Lyon étaient donc de futurs professeurs de français langue étrangère en formation professionnelle. Ils avaient choisi de suivre ce module de formation qui leur permettait de s'entrainer à enseigner le français et à utiliser des outils technologiques à des fins pédagogiques. Nous les appelons désormais les *apprentis-enseignants*. Les étudiants à DCU étaient inscrits dans une école de commerce et étaient en troisième semestre d'apprentissage du français. Ils apprenaient le français avec Françoise Blin en vue d'effectuer un stage professionnelle en France l'année suivante. Nous les appelons désormais les *apprenants*.

1.3.1.1 Les participants

Pour des raisons éthiques qui seront discutées dans le chapitre 4, nous avons conservé pour notre étude sept apprentis-enseignants (cinq femmes et deux hommes) et douze apprenants (huit femmes et quatre hommes) (voire Figure 8). Les apprentis-enseignants avaient environ 22 ans sauf une femme plus âgée en reprise d'études. Ils étaient tous de

[5] Interactions, Corpus, Apprentissages, Représentations.

langue maternelle française, et un apprenti-enseignant était d'origine québécoise. Ils avaient peu ou pas d'expérience d'enseignement. Les apprenants étaient plus jeunes, toujours en licence. Ils étaient tous de niveau B1-B2 en français selon le CECR. Les apprentis-enseignants ont été appariés de manière aléatoire avec les apprenants, un ou deux apprenants par apprenti-enseignant. Sauf en cas d'absence, les binômes et trinômes restaient inchangés durant le semestre.

1.3.1.2 Le scénario pédagogique

Une fois par semaine, les apprentis-enseignants interagissaient avec les apprenants par visioconférence pendant 25-40 minutes. En amont des interactions, les apprentis-enseignants planifiaient de manière collaborative les séances en prenant en compte les besoins langagiers des apprenants et leurs besoins professionnels associés. Les apprenants étant en formation commerciale, les apprentis-enseignants choisissaient pour les activités des thèmes relatifs au monde du travail en France (voir Figure 8).

1.3.2 Le dispositif technique

Les lieux physiques où se sont déroulées les interactions sont des salles informatiques classiques dans lesquelles y a des bureaux, des chaises, des ordinateurs, des micro-casques et des webcams. La photo 1 de la Figure 4 ci-dessous représente la salle lyonnaise avec le projecteur au fond qui a été utilisé pour la conception en groupe des séances. La photo 2 représente la salle dublinoise. La photo 3 représente Adèle[6], une des apprentis-enseignants, à son poste, en train d'utiliser son propre ordinateur portable pour l'interaction. Certains apprentis-enseignants ont décidé d'utiliser leurs propres ordinateurs car ils sont plus puissants que les ordinateurs de bureau installés à l'université.



Figure 4 : Apprentis-enseignants en salle informatique à Lyon 2 et à DCU

Comme le rappelle Peterson (2010), Il faut prendre en compte les spécificités du dispositif et de la technologie quand on analyse des interactions pédagogiques en ligne. Les

[6] Des pseudonymes sont utilisés pour nommer les participants.

interactions qui font l'objet de notre étude ont été médiatisées par Visu (Bétrancourt, Guichon, & Prié, 2011 ; Guichon, Bétrancourt, & Prié, 2012), plateforme qui a été conçue par des informaticiens, des développeurs et des chercheurs en informatique et en didactique des langues de l'Université de Lyon pour l'enseignement synchrone des langues à distance. Visu dispose d'un salon synchrone, un salon de rétrospection et un éditeur de bilans. Nous allons décrire rapidement le fonctionnement de ces outils afin que le lecteur puisse mieux comprendre la façon dont ces interactions ont été médiatisées.

1.3.2.1 Le salon synchrone de Visu

Comme Skype, qui est devenu un métonyme des plateformes de visioconférence, le salon synchrone de Visu dispose des outils de base : un canal audio, un canal vidéo, et un canal clavardage. Dans Visu, les tailles des fenêtres des interactants peuvent être réglées par l'utilisateur (1). Puisque Visu est optimisé pour l'enseignement des langues, il dispose de plus d'outils que les plateformes génériques comme Skype. Certains de ces outils permettent à l'enseignant de faire afficher sur l'écran de l'apprenant des éléments nécessaires à l'activité pédagogique en cours, tandis que d'autres outils existent exclusivement pour l'enseignant et sont invisibles pour l'apprenant. Nous allons décrire ces deux catégories d'outils.

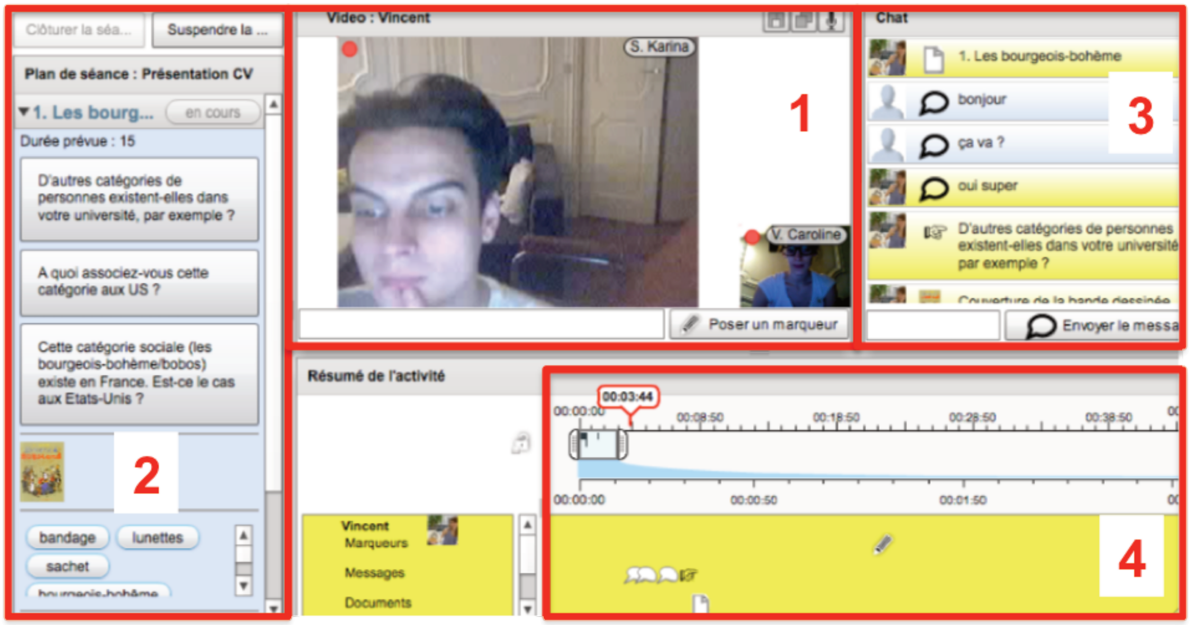


Figure 5 : Le salon synchrone de Visu du point de vue de l'enseignant

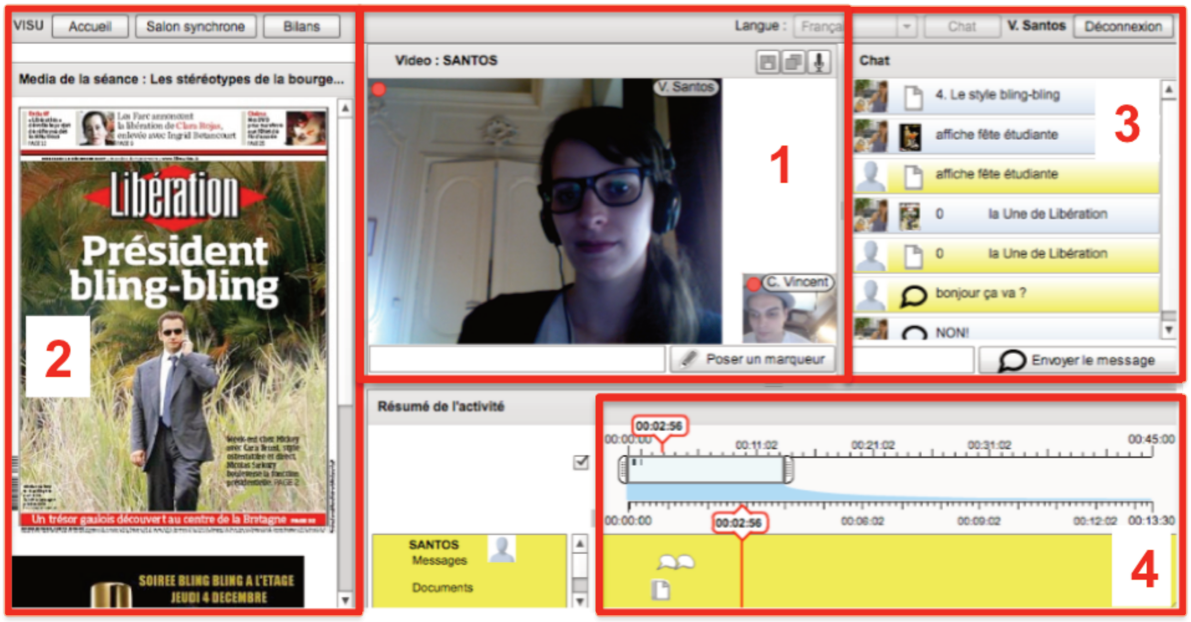


Figure 6 : Le salon synchrone de Visu du point de vue de l'apprenant

Le plan de la séance (voir Figure 5) est créé en amont de l'interaction et s'affiche en permanence sur la gauche de l'écran de l'enseignant (2). Il permet à celui-ci de gérer les activités et de faire apparaître à gauche de l'écran de l'apprenant (2) (voir Figure 6) des documents audiovisuels divers ainsi que des consignes et mots-clés qui s'affichent dans la fenêtre du clavardage (3). Ces consignes, mots-clés et documents audiovisuels doivent être créés et ajoutés au dispositif en amont de l'interaction. Parmi les outils privés, l'enseignant dispose d'aide-mémoires (en dessous des mots-clefs (2), invisible sur la capture d'écran) qui ne s'affichent que sur son propre écran. Ceux-ci permettent aux apprentis-enseignants de créer des memos qui restent accessibles lors de l'interaction. Au cours de l'interaction, l'enseignant peut également poser des marqueurs qui s'affichent sur la ligne de temps (4). Ceux-ci sont invisibles pour les apprenants et permettent à l'enseignant de se référer ultérieurement sans interrompre la conversation. Visible pour tous les interlocuteurs, la ligne de temps est une ligne horizontale qui indique le temps écoulé et précise les moments où les marqueurs ont été posés et où les messages, consignes et documents ont été envoyés

1.3.2.2 Le salon de rétrospection de Visu

Toutes les interactions ont été enregistrées par un logiciel de capture d'écran dynamique. Cet outil permet d'enregistrer tout ce qui se passe sur l'écran de l'ordinateur ainsi que le son. Cet outil a l'avantage d'être peu intrusif (Guichon, 2013). En effet, chaque capture dynamique produit un fichier vidéo que l'on peut regarder ultérieurement. C'est grâce à ces captures d'écran que le salon de rétrospection peut fonctionner. Grâce à ce salon, l'enseignant

ainsi que le chercheur/observateur a accès à la totalité de l'interaction audiovisuelle ainsi que des marqueurs, consignes, mots-clés et documents qui ont été posés ou envoyés au cours de l'interaction (Guichon, 2013) (voir Figure 7).

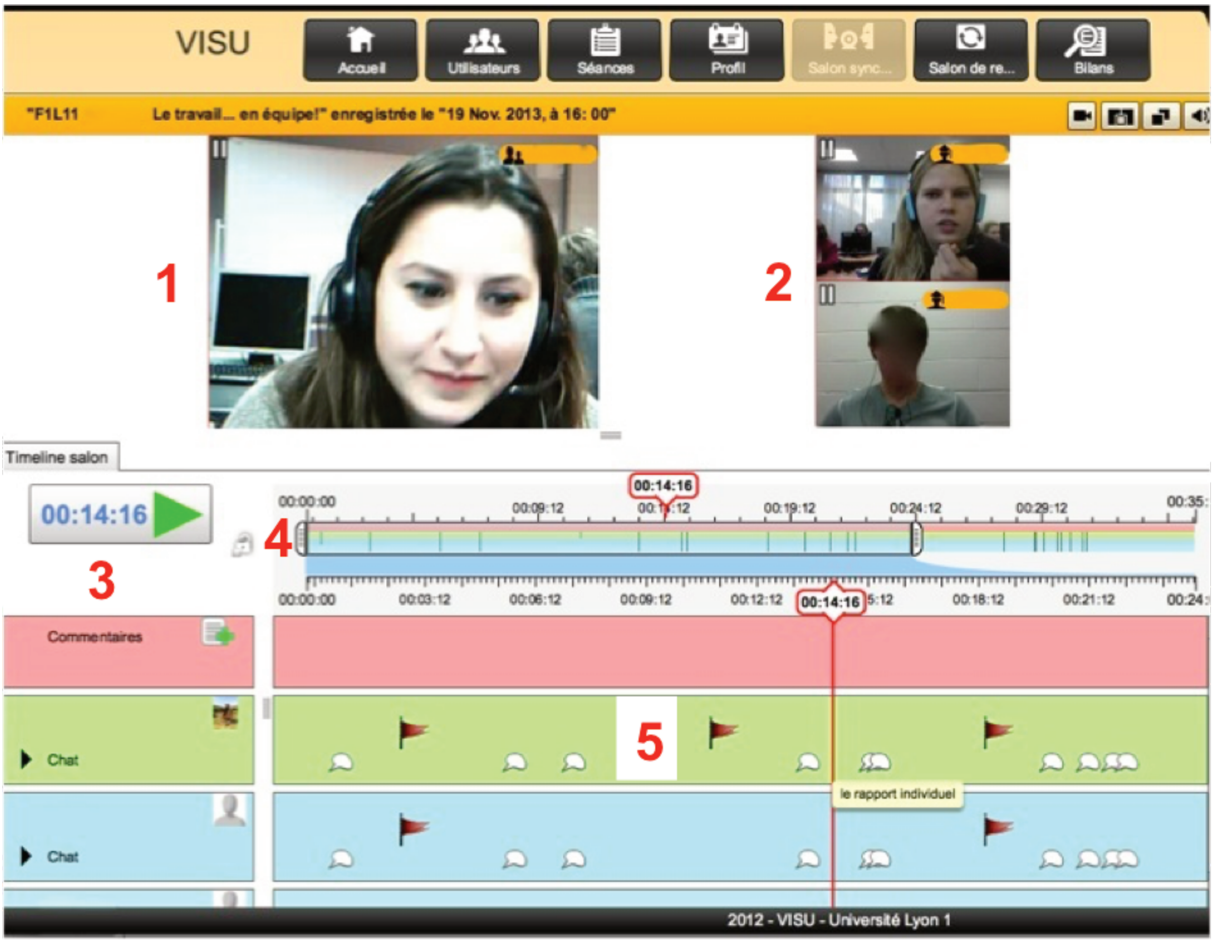


Figure 7 : Le salon de rétrospection de Visu

De même que dans le salon synchrone de Visu, le salon de rétrospection nous permet de voir l'image de l'enseignant (1) et celles des apprenants (2). L'observateur peut régler la taille des fenêtres des interactants de façon à ce que l'image d'un participant soit deux fois plus grande que celle des autres. En cliquant sur « lecture » (3), l'interaction se déroule comme si elle se déroulait en temps réel. La ligne de temps (4) est interactive dans la mesure où on peut zoomer et dézoomer. Ainsi, l'échelle temporelle est modifiable et dynamique. En dessous de la ligne temporelle (5), les consignes, activités, documents et messages envoyés sont visibles quand on met le curseur dessus. En l'occurrence, le curseur est sur un des messages qui ont été envoyés par clavardage dont le contenu est « le rapport individuel ».

Cet outil de grande importance a été développé afin que les apprentis-enseignants puissent avoir un regard critique sur leur propre activité. Ces derniers, dans le cadre de leur formation, s'en servaient pour repérer des moments potentiellement acquisitionnels. Ces

moments ont été ensuite discutés lors des débriefings le lendemain des interactions. Ces débriefings ont été enregistrés par Cathy Cohen. Guichon (2013 : 111) explique que « la rétrospection fournit l'opportunité de mettre en relief les compétences professionnelles par le biais de ce [qu'il] propose d'appeler une **conscience sémio-pédagogique critique** ».

Cet outil s'est révélé également indispensable pour notre collecte de données. C'est dans le salon de rétrospection que nous avons eu notre premier contact avec le corpus. Ce salon nous a permis de sélectionner les interactions à observer en fonction de leur disponibilité. Après avoir pris connaissance des données dans le salon de rétrospection, nous avons procédé à la construction du corpus (cf. chapitre 4).

1.3.3 Le corpus

Après avoir sélectionné les participants, nous avons, en collaboration avec des membres du projet ISMAEL, l'équipe CCC[7] du laboratoire ICAR, et quelques étudiants en Master, construit et transcrit vingt-huit séances. Le tableau ci-dessous donne une vue synoptique du corpus et mentionne notamment les pseudonymes des participants, les thèmes des séances, les durées de celles-ci, et la présence ou non d'une séance de débriefing.

Apprentis-enseignants	Apprenants		Séance 1 15 octobre Les droits du travail	Séance 2 22 octobre Expériences professionnelles	Séance 3 12 novembre Se préparer pour un stage à Reims	Séance 4 19 novembre Gestion d'un projet	Séance 5 26 novembre Mettre en place un projet	Séance 6 3 décembre Entretien d'embauche
Adèle*	Alannah	Catriona*	☑ 43m48s	☑ 40m55s	☑ 33m03s	☑ 34m27s	☑ 41m04s	☑ 35m55s
Emilie*	Aiden	Fiona	☑ 22m50s	☑ 33m00s		☑ 25m37s		☑ 24m12s
Samia*	Sean*	Angela*	☑ 30m20s	Sean absent ☑ 34m34s	⊙ ☑ 31m50s	☑ 35m20s	☑ 39m46s	x
Etienne*	Conor	Sophie				☑ 20m57s		
Mélissa*	Ana*	Alejandra	☑ 42m11s	☑ 39m24s	☑ 37m03s	☑ 30m13s	☑ 32m28s	☑ 35m53s
Séverine*	Naomi			x		☑ 34m47s		
Victor*	Liam		☑ 33m25s	Séance avec Conor ☑ 27m46s	☑ 29m18s	☑ 34m00s	⊙ ☑ 29m01s	x

☑ Séance transcrite
x Séance non disponible
* Entretien
⊙ Vue du hors champ disponible
Temps total transcrit : 15h, 23m, 07s
Planning de la séance
Débriefing le lendemain

Figure 8 : Vue synoptique du corpus ISMAEL

[7] Cellule Corpus Complexes (cf. chapitre 4).

Les sept apprentis-enseignants figurent sur la colonne de gauche, et les apprenants sont représentés à droite. Les images de certains apprenants ont été anonymisées pour des raisons éthiques. Tous les enjeux techniques et éthiques relatifs à la construction du corpus seront expliqués en détail dans le chapitre 4. Ce corpus, constituée de vingt-huit séances et plus de quinze heures d'enregistrement vidéo, nous permet d'aborder plusieurs questions didactiques et scientifiques.

1.3.3.1 Utilité scientifique du corpus

Ce corpus nous donne la possibilité de traiter des questions didactiques. Nous avons évoqué plus haut qu'en visioconférence, les apprentis-enseignants doivent apprendre à gérer le passage entre la focalisation sur la communication médiée, la focalisation sur la tâche, et la focalisation sur la langue, ce que nous avons nommé trifocalisation. Ces moments, où un changement d'orientation s'observe, et où la langue devient un objet de focalisation, sont dignes d'intérêt grâce à leur potentiel acquisitionnel (de Pietro, Matthey, & Py, 1989). Ces moments sont souvent appelés des *séquences de négociation du sens*.

Sarré (2011), en contexte d'apprentissage d'anglais à distance, a comparé la négociation du sens par forum, par clavardage, et par visioconférence. Il a démontré que la visioconférence est un contexte propice à l'émergence de séquences de négociation du sens, car cette modalité de communication a suscité plus de séquences de négociation que le clavardage et le forum. Ce constat rejoint celui de Wang (2006). En effet, la visioconférence permet à des participants de modifier leur langage afin de négocier le sens. Nicolaev (2010) a trouvé aussi que les tâches en visioconférence déclenchent des séquences de négociation. En effet, cette auteure a trouvé que la plupart des séquences de négociation du sens sont déclenchées par un item lexical. Par conséquent, selon elle, la négociation du sens en ligne est plus propice à l'acquisition du lexique qu'au développement de compétences morphosyntaxiques. Le constat que le plus grand nombre de séquences de négociation du sens sont déclenchées par un item lexical a été mis en évidence par plusieurs chercheurs (Blake, 2000 ; de Pietro, Matthey, & Py, 1989 ; Ellis, Tanaka, & Yamazaki, 1994 ; Fasel Lauzon, 2014 ; McDonough & Sunitham, 2009 ; Pica, 1994 ; Smith, 2003 ; Williams, 1999). Outre le fait que la plupart des séquences de négociation sont déclenchées par un item lexical, le vocabulaire « demeure une des entrées majeures dans la langue étrangère » (Cuq, 2004 : 64). Selon cet auteur (2004 : 62),

> « L'entrée par le vocabulaire est un penchant naturel à tout apprenant de langue étrangère, car le vocabulaire est pour l'apprenant le canal le plus direct qui le relie à son système conceptuel : il n'y a donc pas de raison de l'en priver ».

La taille conséquente du corpus dont nous disposons nous permet de repérer de nombreuses séquences d'explication lexicale et d'analyser la façon dont celles-ci sont orchestrées multimodalement par les apprentis-enseignants. Nous pouvons donc poser la question suivante : **combien de séquences d'explication lexicale y a-t-il sur les vingt-huit séances transcrites ?** Bien que l'objectif du semestre ne soit pas l'acquisition du vocabulaire, celui-ci fait partie de tout apprentissage langagier et nous paraît être une entrée pertinente.

Puisque nos échanges se déroulent en visioconférence, nos données nous permettent de mesurer l'apport de chaque canal[8] de Visu lors des séquences d'explication lexicale, en particulier la webcam. Plusieurs chercheurs se sont attachés à mesurer l'apport de la visioconférence par rapport à d'autres technologies. Concernant la présence à distance, Wang (2004) a trouvé que la communication par visioconférence a été jugée par ses participants comme plus authentique que la communication par téléphone. Dans son étude, l'apport de l'image a été beaucoup apprécié par les participants et cette auteure soutient que la webcam est un outil de grande valeur pour l'enseignement à distance. Cependant, Guichon & Cohen (2014) ont comparé l'audioconférence et la vidéoconférence en demandant à des étudiants de décrire quatre photos à un interlocuteur à distance. Selon ces chercheurs, le sentiment de présence à distance n'a pas été engendré par la visioconférence plus que par l'audioconférence. Il semble toutefois que la présence de l'image de l'autre sur un écran améliore la compréhension langagière (Sueyoshi & Hardison, 2005).

Bien que nous ne disposions pas de données audiosynchrones pour mener ce genre de comparaison, nous pouvons tout de même analyser la façon dont la webcam est employée par les apprentis-enseignants pour gérer les séquences d'explication lexicale. Dans une approche multimodale, qui sera détaillée dans le chapitre 3, chaque ressource sémiotique, y compris la webcam, contribue à la production du sens. En choisissant comme point d'entrée les explications lexicales, nous pouvons répondre à la question suivante : **comment les canaux de Visu sont-ils utilisés par les apprentis-enseignants lors des séquences d'explication lexicale ?**

Nos données nous permettent d'observer plusieurs apprentis-enseignants dans des conditions très similaires à qui on a donné les mêmes consignes. Bien que les conditions ne soient pas expérimentales, cette particularité permet de mettre en relief différents profils multimodaux. Chaque participant peut s'approprier les ressources selon sa façon. Nous pouvons, grâce au fait que tous les apprentis-enseignants suivent les mêmes plannings de

[8] Le microphone, la webcam, le clavardage, et les documents audiovisuels envoyés à travers la plateforme.

cours, analyser la façon dont cinq apprentis-enseignants expliquent un même item lexical, ce qui est une opportunité rare. Une telle analyse est proposée dans le chapitre 9.

Notre corpus nous permet non seulement d'observer sept apprentis-enseignants qui suivent le même planning de cours (voir la séance 4 de la Figure 8), mais aussi de suivre les mêmes apprentis-enseignants tout au long du semestre. Ceci offre la possibilité de repérer la présence d'éventuelles évolutions concernant les pratiques multimodales. Nous pouvons donc répondre à la question suivante : **du début à la fin du semestre, y a-t-il des changements remarquables quant à l'usage des canaux de Visu lors des explications lexicales ?** A cette question s'ajoute une autre question : **un semestre de six séances est-il suffisamment long pour mener une étude longitudinale ?**

Notre corpus est riche en raison non seulement de sa taille, mais aussi de la nature des données. Celles-ci sont audiovisuelles, ce qui permet donc des analyses multimodales (cf. chapitre 3). Elles ont été enregistrées par un capteur d'écran dynamique, ce qui permet de capturer de manière peu intrusive les interactions sans perturber et modifier le comportement des interactants.

1.4 Conclusion

Les outils de communication employés par les enseignants de langues étrangères suivent les innovations technologiques disponibles au grand public. Puisque la visioconférence est librement disponible et gratuite depuis une quinzaine d'années, il est normal que les enseignants de langue se soient appropriés cet outil. Ce mode de communication permet à des apprenants d'interagir avec des locuteurs natifs de la langue cible sans se déplacer dans un pays lointain, et permet à de futurs enseignants de langues étrangères en formation professionnelle de s'entraîner. Ce type de dispositif a fait l'objet de plusieurs projets de recherche, notamment le F1L et le projet ISMAEL. Les études menées sur l'enseignement par visioconférence ont démontré que c'est un contexte propice à la négociation du sens. Elles ont montré en outre que la plupart des séquences de négociation sont déclenchées par un item lexical. C'est la raison principale pour laquelle nous nous avons décidé de nous focaliser sur les séquences d'explication lexicale. L'autre raison est que l'enseignement explicite du vocabulaire est nécessaire, comme nous le verrons dans le prochain chapitre.

Nos données sont issues du premier semestre de l'année universitaire 2013-2014 où de futurs professeurs de FLE à Lyon enseignaient le français par visioconférence à des

apprenants à Dublin. Toutes ces interactions ont été enregistrées. En collaboration avec plusieurs personnes d'équipes différentes, nous avons construit un corpus de vingt-huit séances entre sept apprentis-enseignants et douze apprenants. Ces données audiovisuelles nous permettent d'analyser, en adoptant une approche multimodale, la façon dont les apprentis-enseignants tirent profit des ressources du dispositif pour orchestrer les explications lexicales. Nos quatre chapitres d'analyse seront consacrés à cette question de recherche centrale. Nos données nous permettent aussi d'aborder la question de la durée nécessaire pour mener une étude longitudinale. Nous tenterons d'apporter une réponse à cette question annexe lors de la discussion qui aura lieu dans le chapitre 10.

Si, à notre connaissance, il n'existe pas d'études qui analysent uniquement les séquences d'explication lexicale pendant un semestre entier d'interactions en visioconférence, il existe déjà un vaste champ de recherche sur l'enseignement et l'apprentissage du vocabulaire en langue étrangère. C'est à ceci que nous consacrerons notre prochain chapitre.

Chapitre 2 : Enseignement et apprentissage du vocabulaire en langue étrangère

2.1 Introduction

Avant d'analyser les séquences d'explication lexicale en visioconférence, nous avons besoin de repères théoriques afin de porter un regard critique sur ce que nous observons. Ce chapitre se donne pour objectif de définir un item lexical et d'esquisser les facteurs susceptibles de favoriser leur acquisition. Dans un premier temps, nous montrerons la nécessité d'un apprentissage explicite du vocabulaire en classe de langue étrangère. Dans un deuxième temps, nous aborderons la notion d'item lexical ainsi que la connaissance de celui-ci. Enfin, en s'appuyant sur la recherche issue des domaines de la psychopédagogie et de la didactique des langues, nous résumerons les pratiques susceptibles de favoriser l'acquisition du vocabulaire et de rendre les séquences d'explication lexicale efficaces.

2.2 Pourquoi privilégier l'enseignement explicite du vocabulaire en classe de langue étrangère ?

Les locuteurs natifs adultes connaissent environ 20 000 familles[1] de mots dans leur langue maternelle (Nation, 2013). Hormis les premiers milliers de mots, classés par ordre de fréquence, ceux-ci sont appris de manière incidente, c'est-à-dire sans se focaliser explicitement sur l'acquisition du vocabulaire (Grossmann, 2011). Bien que l'apprentissage incident apporte des avantages pour l'acquisition d'une langue étrangère — les mots appris de cette façon sont forcément contextualisés, on peut se concentrer à la fois sur la langue et le contenu en lisant, et l'apprentissage est plus individualisé quand l'apprenant choisit ses propres textes à lire (Huckin & Coady, 1999) — cette façon d'apprendre du vocabulaire en langue étrangère est problématique pour les raisons suivantes (Hulstijn, Hollander, & Greidanus, 1996) :

- En rencontrant un mot inconnu, l'apprenant risque de ne pas le repérer ou de croire à tort qu'il le connaît ;

[1] Une famille de mots inclut toutes les formes dérivées d'un mot (masculin, féminin, pluriel, formes fléchies, etc.).

- Même si l'apprenant repère un mot inconnu, il peut décider de passer outre ;
- Le contexte peut être tellement redondant[2] que l'apprenant ne fait pas le lien entre la forme du mot et son sens ;
- Souvent, le sens du mot ne peut pas être inféré en s'appuyant sur le contexte, et l'apprenant risque de deviner incorrectement le sens du mot ;
- Même si l'apprenant dispose d'un dictionnaire, il ne le consulte pas toujours ;
- Voir une seule fois un mot ne suffit pas pour l'apprendre, ce qui fait que l'apprentissage incident est un processus extrêmement lent.[3]

Une autre raison pour laquelle l'apprentissage incident du vocabulaire en langue étrangère est problématique est que cette méthode d'apprentissage ne fonctionne que pour les apprenants les plus avancés. Apprendre du vocabulaire en lisant ou en écoutant en s'appuyant sur le contexte nécessite que l'on connaisse déjà beaucoup de mots. Pour comprendre un texte, du moins en anglais, il faut connaître à peu près 95% des mots, soit environ 3 000 à 4 000 mots (Nation, 2013). A ce niveau, un mot sur vingt sera inconnu. Or, pour être en mesure d'inférer le sens des mots inconnus, il faut comprendre 98% des mots du texte (Huckin & Coady, 1999), ce qui nécessite un vocabulaire encore plus riche, de l'ordre de 6 000 à 9 000 mots. Schmitt (2008) estime que pour comprendre 98% d'un discours oral, il faut connaître entre 5 000 et 7 000 mots. En-dessous de ce seuil, l'apprentissage incident de nouveaux mots est problématique car il n'y a pas assez de mots connus pour donner le contexte nécessaire pour inférer le sens des mots inconnus. Ceci est valable aussi pour l'oral : Bonk (2000) a trouvé une corrélation significative et positive entre reconnaissance lexicale et compréhension de textes oraux. Ainsi, pour apprendre de nouveaux mots, il faut en connaître déjà beaucoup. Folse (2004) associe ce phénomène à l'Effet Matthieu.[4]

Puisque la langue ne s'apprend pas uniquement par absorption (Théophanous, 2001), il faut se livrer à un enseignement explicite et systématique du vocabulaire afin d'apprendre au plus vite les milliers de mots qui sont nécessaires pour une compréhension suffisante de textes et pour être autonome. Cette grande tâche d'apprentissage s'impose davantage pour les apprenants souhaitant poursuivre des études universitaires dans une langue étrangère, où la compréhension de textes de première année nécessite une connaissance d'environ 10 000

[2] Coady (1993) explique que si le contexte est trop redondant, le lecteur n'aura pas besoin de connaître le mot pour comprendre le texte.

[3] Hill & Laufer (2003) estiment qu'un apprenant peut espérer apprendre 3-4 mots par 20 000 mots lus, ce qui fait qu'il faudrait lire 420 romans pour apprendre 2000 mots.

[4] L'Effet Matthieu désigne de manière très générale les mécanismes par lesquels les plus favorisés tendent à accroître leur avantage sur les autres.

mots[5]. Le temps d'apprentissage doit être rentabilisé et optimisé par l'apprentissage direct du vocabulaire en classe de langue (Folse, 2004). Il revient aux enseignants de langues d'enseigner les mots de haute fréquence ainsi que les mots spécifiques du domaine d'intérêt des apprenants, par exemple des mots susceptibles d'être rencontrés lors d'un stage en France.

2.3 Qu'est-ce qu'un item lexical ?

Si l'importance des mots paraît évidente, la définition d'un *mot* est plus complexe en raison de ses multiples façons d'être appréhendé :

> « Unité graphique pour le statisticien, unité rythmique pour le phonéticien, unité de sens (lexème ou lexie) pour le sémanticien, unité minimale de discours (de sens et de fonction) pour le lexicographe et unité d'apprentissage pour le pédagogue qui envisage une approche lexicale » (Tréville, 2000 : 27).

Parmi ces différents chercheurs, certains font la distinction entre *mot graphique* et *mot linguistique*. Mortureux (2008 : 8) explique qu'un mot graphique est « une suite de caractères délimitée par deux espaces », tandis qu'un mot linguistique peut être composé de plusieurs mots graphiques comme par exemple « pommes de terre ». Un seul mot graphique peut renvoyer à plusieurs mots linguistiques : le verbe « voler », « puisque le mot graphique *vol* peut correspondre à deux mots linguistiques différents (homonymes homographes) » (Mortureux, 2008 : 8).

Ensuite, se pose la question d'homonymie et de polysémie, puisque la plupart des signifiants, qu'ils soient composés d'un ou de plusieurs mots graphiques, renvoient à plusieurs signifiés. En effet, les mots changent de sens en fonction du contexte. Beaucoup de mots ont par exemple un sens courant et un sens scientifique[6] (Calaque, 2000). La distinction entre homonymie et polysémie demeure opaque. La différence entre l'homonymie et la polysémie est une question de degré de différence entre les acceptions d'un mot (Mortureux, 2008). Dans un dictionnaire, un mot polysémique aura plusieurs définitions sous la même entrée alors que deux mots homonymes auront deux entrées distinctes. Cette relation est interprétée différemment selon les dictionnaires. Par exemple, le mot *comprendre* est un mot polysémique dans certains dictionnaires (une seule entrée avec plusieurs définitions) tandis que d'autres dictionnaires lui accordent plusieurs entrées (homophonie) (Mortureux, 2008). Bogaards (1994 : 17) pense qu'il n'y a pas de rupture nette entre homophonie et polysémie : « Qu'il y ait ou non une opposition nette entre homonymes et mots polysémiques, ou que ces

[5] Étude faite par Hazenberg & Hulstijn (1996) concernant l'apprentissage du néerlandais.

[6] Par exemple, le mot « poids », scientifiquement parlant, est la force qui attire les objets vers la Terre. Au sens courant, le mot « poids » se confond avec le mot « masse », qui veut dire la quantité de matière.

deux catégories de mots constituent plutôt les points extrêmes d'un continuum, cela ne fait pas beaucoup de différence pour l'apprenant d'une L2 ».

Qu'il s'agisse de polysémie ou d'homophonie, un mot hors contexte peut avoir plusieurs sens différents. En revanche, dans un discours donné, un mot est toujours monoréférentiel sauf dans certains cas exceptionnels comme les jeux de mots. Lors de la production et de la réception du langage, les interlocuteurs doivent sélectionner la bonne acception du mot. Ceci renvoie à la distinction entre *vocable* et *lexème*. Le vocable est ce que l'on observe dans un discours, et le lexème est une réalité abstraite. Un lexique, qui n'est qu'une hypothèse faite par des linguistes, est l' « ensemble des lexèmes pourvus d'une signification abstraite » (Mortureux, 2008 : 11). Dans l'interaction, la compétence lexicale est actualisée lors de l'énonciation quand les lexèmes sont convertis en vocables.

Afin de résoudre la complexité associée aux phénomènes brièvement abordés ci-dessus (mot graphique, mot linguistique, homophonie, polysémie, vocable, lexème), Bogaards (1994 : 67) propose le concept d'*unité lexicale* qui « unit une forme bien déterminée et un sens stable ». L'unité lexicale doit être au moins un mot et doit être au moins un constituant sémantique. Ainsi, *mé* dans *mécontant*, *méfiance* et *mévente*, n'est pas une unité lexicale, bien qu'il soit un constituant sémantique, car il n'est pas un mot. De même, *lune* dans *lune de miel* n'est pas une unité lexicale, bien qu'il soit un mot, car il n'est pas un constituant sémantique dans cette construction. Une unité lexicale, telle qu'elle a été définie par Bogaards, peut être un mot graphique, un mot linguistique, ou une expression idiomatique. C'est le mariage d'une forme et d'un sens. Selon cet auteur, il est plus judicieux d'un point de vue didactique d'isoler l'unité lexicale en question lors d'une séquence d'explication.

Une unité lexicale telle qu'elle est décrite par Bogaards ne correspond qu'à une seule acception d'un mot. Ainsi, pour citer un exemple que donne cet auteur, le mot « rayon » a neuf sens qui correspondent à neuf unités lexicales différentes. Selon nous, conceptualiser les différentes acceptions du mot « rayon » comme étant neuf unités lexicales différentes ne constitue pas une bonne stratégie didactique. La capacité à sélectionner la bonne acception d'un mot est une stratégie qui devrait être enseignée, ainsi que la capacité à trouver le sens commun de toutes les acceptions (Nation, 2013). Cela implique que « rayon » soit considéré comme un seul mot avec neuf acceptions. Nous privilégions avec Nation (2013) l'approche polysémique[7] qui apporte plusieurs avantages. Notamment, à chaque fois que l'on rencontre une de ces acceptions, celle-ci sert de répétition en renforçant la connaissance du mot.

[7] Dans une approche polysémique, « rayon » est un seul item lexical ayant neuf sens différents.

L'apprentissage de nouvelles acceptions d'un mot déjà rencontré consiste à élargir les connaissances actuelles au lieu d'ajouter une toute nouvelle unité lexicale à son stock de connaissances lexicales. Enfin, reconnaître ou utiliser une acception d'un mot qui est autre que celle initialement apprise constitue ce que Joe (1995, 1998) nomme l'usage génératif, favorisant le traitement profond et la mémorisation de mots. Le concept d'usage génératif n'est pas possible si chaque acception d'un mot renvoie à une unité lexicale différente. Pour ces raisons, nous privilégions les termes *mot* et *item lexical* plutôt que le terme *unité lexicale*.

Certains items lexicaux sont composés de plusieurs mots graphiques. Ceux-ci ont plusieurs noms. (Théophanous, 2008 : 47) en donne 18 :

Collocations	Expressions idiomatiques	Locutions
Expressions stéréotypées	Expressions figées	Holophrases
Formules	Formulations	Phrasèmes
Groupements phraséologiques	Idiotismes	Idiomatismes
Routines discursives	Phrasillons	Pragmatèmes
Tournures	Clichés	Lieux communs

Tréville et Duquette (1996 : 16) expliquent que ces locutions sont figées ou semi-figées « suivant que la soudure entre les morphèmes est plus ou moins forte et plus ou moins stable ». Les unités polylexicales varient aussi selon le degré d'idiotisme. C'est-à-dire que le sens de certaines unités polylexicales peut être deviné si on connaît le sens des mots graphiques qui les composent, tandis que d'autres nécessitent un apprentissage explicite. Nous considérons que *collocation* est une unité polylexicale dont le sens peut être inféré si on connaît le sens des mots graphiques qui la compose, tandis qu'une *expression idiomatique* est une unité polylexicale dont le sens ne peut pas être inféré en regardant les composants. Celles-ci sont donc à apprendre comme des mots car elles ont une forme unique qui renvoie à un sens qui ne peut pas être inféré en regardant les composants. Bogaards (1994 : 20) donne quelques exemples :

> « *mettre la main à la pâte* ou *avoir un chat dans la gorge*, où la connaissance de tous les mots ne garantit pas une interprétation correcte. Malgré leur forme complexe, les expressions idiomatiques sont simples au point de vue sémantique, c'est-à-dire que la signification de l'ensemble ne se laisse pas inférer des significations individuelles des éléments composants ».

Tréville et Duquette sont du même avis de Bogaards. Ces expressions idiomatiques doivent être traitées comme des mots car elles relient une forme et un sens. Elles doivent être enseignées explicitement car leur apprentissage en milieu naturel est difficile, même pour les apprenants les plus avancés (Huckin & Coady, 1999). Une source de difficulté est que les expressions idiomatiques ne sont pas systématiquement « balisées à l'aide de marqueurs tels que des guillemets à l'écrit ou des mises en relief intonatives à l'oral » (Tréville & Duquette, 1996 : 17). Les mots graphiques, les mots linguistiques, les collocations, et les expressions

idiomatiques sont tous des items lexicaux à apprendre. Le tableau ci-dessous exemplifie chacun de ces termes.

Terme	Exemple
Mot graphique	manageur
Mot linguistique	agent de maintenance
Collocation	moment de détente
Expression idiomatique	arrondir ses fins de mois

Figure 1 : Différents types d'items lexicaux

Dans la section suivante, nous allons explorer la complexité de la connaissance d'un item lexical.

2.3.1 Connaître un item lexical

Toute explication lexicale a pour objectif de transférer des connaissances, et nos analyses didactiques ont pour objectif d'étudier ce transfert de connaissances. Afin de mener de telles analyses, il faut décomposer le concept de connaissance d'un item lexical. Certains auteurs conçoivent la connaissance d'un item lexical par familiarité, et d'autres par facettes. Les connaissances productives et réceptives sont souvent distinguées.

La métaphore classique de la connaissance d'un mot est celle d'une médaille à deux côtés : la forme d'un côté et le sens de l'autre (Van der Linden, 2006). Le mot est considéré comme un signe lexical, et chaque signe possède un signifiant (forme) et un signifié (sens). Selon Mortureux (2008 : 21), « les deux sont aussi indissociables que les deux faces d'une feuille de papier ». Ce signe a la capacité de se référer à des éléments de la réalité, mais « Cette ***référence*** reste ***virtuelle*** tant qu'elle ne se trouve pas ***actualisée*** par l'emploi du mot en discours » (Mortureux, 2008 : 21). Connaître un mot (signe) consiste donc à connaître ces deux dimensions (signifiant et signifié). La compétence productive consiste à être capable d'aller du sens vers la forme (du signifié vers le signifiant), et la compétence réceptive consiste, à l'inverse, à être capable d'aller de la forme vers le sens (du signifiant vers le signifié).

La connaissance d'un mot est aussi conçue comme étant une échelle allant d'une faible connaissance à une forte connaissance. Au niveau zéro, le mot n'a jamais été rencontré et est complètement inconnu. Au niveau maximum, on est capable d'en donner une définition et de l'employer de manière appropriée dans une phrase originale (Waring, 2002). Au fil de l'apprentissage, la connaissance d'un mot progresserait sur cette échelle linéaire. Tréville &

Duquette (1996) conçoivent la connaissance d'un mot comme étant une échelle allant progressivement d'une connaissance réceptive à une connaissance productive. Waring critique cette conception linéaire de la connaissance d'un mot, car les connaissances réceptives et productives ne sont pas nécessairement liées. Il propose de concevoir la connaissance d'un mot par états, où les connaissances réceptives et productives sont distinctes.

Certains auteurs vont au-delà de la conception classique (forme et sens) évoquée ci-dessus et considèrent la connaissance d'un mot comme étant pluridimensionnelle. Pour Laufer (1994 : 2)[8], il y aurait six dimensions de la connaissance d'un mot :

> « (a) Forme : reconnaissance de la forme orale et écrire ; capacité à prononcer et à orthographier le mot correctement ;
>
> (b) Structure lexicale : reconnaissance du radical et des affixes ; capacité à produire des dérivés du mot ;
>
> (c) syntaxe du mot dans un syntagme / dans une phrase ;
>
> (d) Sens : référentiel (polysémie et usages métaphoriques éventuels), affectif (ensemble des connotations), pragmatique (adéquation en situation) ;
>
> (e) Relations lexicales du mot avec d'autres mots : synonymie, antonymie, hyponymie, etc.
>
> (f) Collocations les plus fréquentes. »[9]

Ces six dimensions montrent la complexité de la connaissance d'un item lexical et à quel point lexique, grammaire et syntaxe s'imbriquent. Nous verrons plus bas que ces six dimensions peuvent être condensées en trois. Pour García (2008 : 13), la connaissance d'un mot serait tridimensionnelle :

> « – d'un point de vue pragmatique, connaître un mot signifie savoir l'utiliser avec sa signification exacte dans son contexte et savoir qu'il est accompagné d'autres éléments non verbaux (gestes, intonation, etc.) ;
>
> – d'un point de vue socioculturel, connaître un mot signifie connaître le registre dans lequel il est utilisé, s'il s'agit d'un mot que l'on utilise dans un dialecte concret ou dans une zone géographique particulière, voire un pays parmi tous les pays de la langue ;
>
> – et d'un point de vue linguistique, connaître un mot signifie pouvoir l'appréhender sur différents plans : sur le plan sémantique, pouvoir comprendre son/ses sens ; sur le plan morphologique et syntaxique, connaître les préfixes et les suffixes avec lesquels on peut l'utiliser, connaître ses dérivés, etc. ; enfin, sur un plan phonétique, pouvoir le prononcer correctement tout comme pouvoir le reconnaître à l'oral. »[10]

Cette conception de la connaissance d'un item lexical montre que savoir l'employer est aussi important que connaître son sens, et à quel point connaissances linguistiques et savoir-faire social sont intimement liés.

---

[8] Nous disposons du manuscrit auteur qui est numéroté de 1 à 14.

[9] Les dimensions *a* et *b* feraient partie de la facette « forme » de Nation (2013). Les dimensions *c* et *f* feraient partie de la facette « usage », et les dimensions *d* et *e* feraient partie de la facette « sens ».

[10] Ces deux catégories tomberaient sur la facette « usage » de Nation (2013) tandis que la troisième serait à cheval entre les facettes que Nation appelle « forme » et « sens ».

Nation (2013) résume la conception de la connaissance d'un mot et regroupe les dimensions en trois grandes catégories : la forme, le sens et l'usage. A l'aide de trois exemples tirés de notre corpus, nous montrons que cette conception de la connaissance d'un item lexical est pertinente pour notre étude.

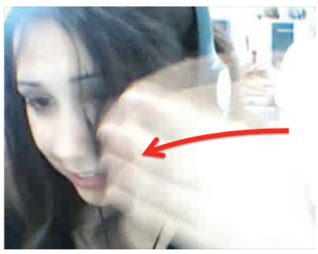
Dans ce premier exemple, Catriona n'arrive pas à produire la forme du mot « employé » en français. Elle semble déjà connaître le sens. Adèle se focalise sur la forme du mot afin d'aider son apprenante à le prononcer.

Transcription de la parole	**Image d'Adèle**	**Phénomènes observés**
1 CTR oui\ et: hum: (0,5) le: (.) em- (0,5) "the work euh the" em- (0,7) l'em- (0,5) empl- (0,4) em- (0,9) empl- (0,5) emploi le em- (0,3) em- 2 ADL **EMPLOYÉ**		Adèle donne oralement la forme du mot « employé » en haussant les sourcils et en rendant visibles ses mouvements de bouche

Exemple 1 : Une focalisation sur la forme d'un item lexical[11]

En se rapprochant de la webcam, en haussant les sourcils, et en élevant sa voix, Adèle donne la forme correcte du mot « employé ». Cette séquence porte uniquement sur la forme de l'item lexical.

Dans l'exemple suivant, Emilie anticipe une incompréhension du sens du mot « extrascolaire » et l'explique à l'aide d'une définition et d'un geste métaphorique.

Transcription de la parole	**Image d'Emilie**	**Phénomènes observés**
1 EMI est-ce que vous avez des activités/ extrascolaires\ (0,6) c'est à dire des activités que vous pratiquez/ (0,4) en dehors (.) des cours (1,1) qu'est-ce que vous faites **(0,7) à côté**		Emilie fait un geste métaphorique devant la webcam voulant dire « à côté »

Exemple 2 : Une focalisation sur le sens d'un item lexical

[11] Notre façon de transcrire et d'annoter les données sera décrite en détail dans la partie méthodologique. Nous écrivons en caractères gras la partie de la transcription qui correspond à l'image dans la deuxième colonne.

Emilie dirige l'attention vers la forme de l'item lexical en utilisant le logonyme « dire » (cf. chapitre 6). Après s'être focalisée sur la forme, Emilie se focalise sur le sens de l'item lexical en donnant une définition accompagnée d'un geste métaphorique.

Enfin, une explication lexicale peut porter non seulement sur la forme et sur le sens d'un item, mais aussi sur les aspects liés à son usage, comme son registre, sa fréquence, son fonctionnement grammatical ou les mots avec lesquels il se colloque. Dans ce dernier exemple, Victor explique le sens de l'expression « du coup » ainsi que sa fréquence et la manière dont elle est employée en France.

Transcription de la parole		
1	VIC	est-ce que tu connais l'expression/ (.) du coup/ (1,5)
2	LIM	euh: (1,0) [du coup/]
3	VIC	[dans la première ph]rase/ (.) j'ai pas d`voiture/ (.) du coup/ (.) je viendrai au boulot en tram (0,3) du coup tu connais ou pas (0,8)
4	LIM	non (.) euh: je ne connais pas
5	VIC	euh s- souvent les les: les français disent euh beaucoup beaucoup ça en fait moi je: je l`dit pas trop/ (0,4) mais en france du coup euh pff (0,4) on l'emploie euh:\ (.) pour dire hum:\ (0,7) alors/ normalement/ ça a un sens de hum: (0,4) de conséquence (0,2) tu vois/ (0,4) j'ai pas d`voiture/ du coup donc ça explique/ (.) j- (0,3) donc (.) comme j'ai pas d`voiture/ (.) je viendrai au boulot en tram\ (0,9) donc c'est c'est une euh relation de cause/ à effet en fait\ (1,8)
6	LIM	ok (0,7) [d'accord]
7	VIC	mais [souvent] les français commencent leurs phrases avec euh du coup/ euh et puis y a aucun lien avec euh c`qu'ils disaient auparavant (0,5) donc c'est un peu un: (1,3) euh: un tic de langage (.) comme euh on peut dire en anglais euh "so"/ euh à toutes les sauces quoi

Exemple 3 : Une focalisation sur l'usage d'un item lexical

Victor dirige l'attention vers la forme de l'expression « du coup » à l'aide du logonyme « expression » (ligne 1), et à l'aide de l'autonymie à la ligne 3 (cf. chapitre 6). Avant de donner une définition de l'expression, Victor parle de la fréquence à laquelle elle est employée en France (« souvent », « beaucoup », ligne 5). A la suite de sa définition, Victor qualifie l'expression d'un « tic de langage » (ligne 7) et tente une comparaison avec un mot anglais qui est susceptible d'être surutilisé.

Ces trois exemples montrent que la connaissance d'un item lexical peut être divisée en trois facettes principales : la forme, le sens, et l'usage. Les chapitres 6 et 7 seront consacrés respectivement sur la focalisation sur la forme et sur le sens des items lexicaux. La focalisation sur l'usage sera abordée au chapitre 9 lors d'une comparaison avec cinq apprentis-enseignants expliquant le même item lexical.

Nation (2013 : 49) subdivise chacune de ces catégories en trois sous-catégories pour donner neuf facettes de la connaissance d'un mot. Les connaissances productive et réceptive sont toujours distinguées :

Forme	Orale	R	A quoi ressemble la forme acoustique ?
		P	Comment prononce-t-on ce mot ?
	Écrite	R	A quoi ressemble la forme écrite ?
		P	Comment écrit-on le mot ?
	Composants	R	Quelles parties sont reconnaissables ?
		P	Quelles sont les parties nécessaires pour exprimer le sens ?
Sens	Forme et sens	R	Quel est le sens de ce mot ?
		P	Quelle forme peut être utilisée pour exprimer ce sens ?
	Concept et référents	R	Qu'est-ce qui est inclus dans le concept ?
		P	A quoi ce concept peut-il faire référence ?
	Associations	R	A quels autres mots ceci nous fait-il penser ?
		P	Quels autres mots peuvent remplacer celui-ci ?
Usage	Fonctions grammaticales	R	Dans quelles structures grammaticales trouve-t-on ce mot ?
		P	Dans quelles structures grammaticales doit-on utiliser ce mot ?
	Collocations	R	Quels sont les autres mots utilisés avec celui-ci ?
		P	Quels autres mots doit-on utiliser avec celui-ci ?
	Contraintes liées à l'usage (registre, fréquence…)	R	Où, quand, et à quelle fréquence rencontre-t-on ce mot ?
		P	Où, quand, et à quelle fréquence peut-on utiliser ce mot ?
R = connaissance réceptive ; P = connaissance productive			

Figure 2 : Les facettes de la connaissance d'un item lexical (Nation, 2013 : 49)

Pour chaque facette, une connaissance réceptive et une connaissance productive y sont associées. Cette distinction est importante car la façon dont un mot est appris a une influence sur le type de connaissance que l'on a de celui-ci. Cela veut dire que si on apprend un mot de manière réceptive, on en aura une connaissance réceptive, et à l'inverse, si on apprend un mot de manière productive, on en aura une connaissance productive[12]. Cependant, les compétences productives et réceptives sont interdépendantes. Nation (2013 : 47) nuance ainsi la distinction classique entre réception et production, car quand on lit ou on écoute, on produit du sens. Il propose donc, pour les tâches classiquement appelées réceptives, « meaning

[12] Par exemple, on acquière une connaissance réceptive d'un item lexical en recherchant son sens dans un dictionnaire. On acquière une connaissance productive d'un item lexical en l'énonçant ou en l'écrivant.

recognition and meaning recall » (la reconnaissance et le rappel du sens), et pour les tâches classiquement appelées productives, « form recognition and form recall » (la reconnaissance et le rappel de la forme). Enfin, si on apprend un mot de manière réceptive on peut espérer le connaître au moins en partie productivement. Si on veut maximiser les connaissances réceptives et productives, l'apprentissage productif est à envisager (Webb, 2005, 2009). La dichotomie entre connaissance réceptive et connaissance productive est aussi nuancée par Tréville (2000 : 53), qui propose la notion de *connaissance optimale* : « Une connaissance optimale suppose que l'on est familiarisé avec tous les aspects du mot : on est capable de définir ses différentes acceptions, on sait l'écrire et l'employer dans le contexte et la situation qui conviennent ».

Nous acceptons cette conception de la connaissance d'un mot en trois grandes facettes : forme, sens et usage. Connaître la forme d'un item lexical implique que l'on soit capable de la lire, l'entendre, la prononcer et l'écrire. Concernant la connaissance du sens d'un item lexical, il faut connaître ce que Bogaards (1994 : 28) nomme le *sens dénotatif* ou *sens cognitif* qui est « le sens qui établit le lien entre la forme de [l'item lexical] et les référents ». On retrouve ce type de sens dans un dictionnaire monolingue ou dans une explication lexicale donnée par un enseignant. Concernant l'usage, il faut savoir à quel registre de langue appartient l'item lexical, sa fréquence, ses conjugaisons, les collocations qui sont formées par l'item, et ses fonctions grammaticales. La grammaire sera abordée dans le paragraphe suivant.

Contrairement à l'esprit d'autrefois, grammairiens et lexicologues se sont rendu compte qu'il n'existe pas de rupture nette entre lexique et grammaire. Ceux-ci sont interdépendants (Binon & Verlinde, 2004; Bogaards, 1994; Cuq, 2004; Folse, 2004; Laufer, 1994; Singleton, 1994; Théophanous, 2008; Tréville, 2000). Cuq (2004 : 61) remarque que « L'originalité du lexique est qu'il est à la fois lié à la grammaire, et justifie donc une certaine focalisation sur la forme, mais aussi à la culture, et justifie donc tout autant une focalisation sur le sens ». Les règles grammaticales « ne fonctionnent pas de façon autonome, mais dépendent du contenu des mots employés » (Bogaards, 1994 : 42). Sans connaissances grammaticales, on peut communiquer, mais sans connaissances lexicales, la communication devient impossible. C'est pourquoi certains disent que le langage consiste en mots grammaticalisés, et non pas en grammaire lexicalisée (Folse, 2004). Chaque mot a des propriétés syntaxiques différentes, ce qui fait que « la syntaxe émane du lexique » (Tréville, 2000 : 37). Si le comportement syntaxique d'un mot en L2 diffère du comportement

syntaxique de son équivalent en L1, cela peut être une source de difficultés, mais uniquement pour la production (Laufer, 1994). Grammaire et lexique étant si intimement liés et indissociables, il ne s'agit pas de se focaliser sur l'apprentissage du lexique au détriment de la grammaire. La connaissance d'un mot implique que l'on soit capable de l'utiliser dans une phrase en respectant les règles grammaticales et syntaxiques qui y sont liées.

Connaître un mot implique que l'on connaisse les autres mots qui sont utilisés en même temps, c'est-à-dire connaître ses collocations. Il faut donc enseigner la façon dont les items lexicaux font partie d'un réseau plus vaste (Polguère, 2004). Comme le rappellent Binon & Verlinde (2004 : 274) « Le lexique ne se compose pas seulement de mots isolés, mais aussi de combinaisons de mots ». Benigno, Grossmann, & Kraif (2015 : 84) rejoignent ces propos en disant que « le bagage lexical d'une langue consiste en une variété de formules, *chunks*, phrases de routine, expressions figées, collocations ». Une collocation, selon Tréville & Duquette (1996), est une locution figée binaire où le verbe et son complément sont reliés. Les co-occurents sont des mots qui n'apparaissent pas forcément ensemble syntaxiquement, mais qui se côtoient dans un même texte. Cette relation n'est pas forcément réciproque. Si on prend l'exemple de « cheveux châtains », châtains se combine presque exclusivement avec cheveux, mais cheveux peut se combiner avec d'autres mots. Ces auteures argumentent que les mots ne prennent leurs sens de manière complète qu'en étant associés avec les mots avec lesquels ils sont souvent en co-occurrence. Les collocations sont différentes selon la langue. Laufer (1994) remarque qu'en anglais on *a* le petit déjeuner (« to have breakfast »), tandis qu'en français on le *prend*. Benigno, Grossmann, & Kraif (2015) critiquent le *Français Fondamental*[13] de Gougenheim car cette liste isolait les mots et ne prenait pas en compte les mots qui y sont souvent associés.

La production de collocations est plus difficile que la compréhension, mais leur maîtrise est avantageuse pour l'apprenant. Tréville (2000 : 106) liste plusieurs avantages liés à l'apprentissage de collocations : «

- elles sont plus fréquentes dans le discours que les unités lexicales simples ;
- elles sont porteuses d'information morphosyntaxique : sur le genre des mots, par exemple (« briser **la** glace », « ronger **son** frein », etc.), sur l'emploi des temps après certaines

[13] Le français fondamental est une « liste de mots et d'indications grammaticales élaborée dans une perspective de diffusion internationale et de démocratisation de l'enseignement du français » élaborée pendant les années 1950 à l'Ecole Normale Supérieure de Saint-Cloud sous la direction de Georges Gougenheim et de Paul Rivenc (Cortier, 2006 : 1). La parution de cette liste est considérée comme une étape majeure dans l'histoire de la didactique du FLE. Selon Cuq (2004 : 108) : « Fondé sur les résultats d'enquêtes statistiques effectuées d'abord (et pour la première fois dans le monde) sur des discours parlés recueillis en situation à l'aide de magnétophones, il a permis d'établir une liste de fréquence du vocabulaire général le plus usuel, ainsi que des termes grammaticaux et de quelques structures syntaxiques très employées dans les échanges oraux ».

conjonctions (« pourvu qu'il **fasse** beau », sur l'usage des comparatifs (« **mieux** vaut tenir **que** courir »), sur l'emploi ou l'absence de prépositions après les verbes (« rien ne sert de courir, il faut partir à point ») ;

- elles permettent à l'apprenant de repérer les comportements collocationnels des mots et d'anticiper la suite d'un groupe de mots comme le font spontanément les locuteurs natifs (« jouer cartes sur… », « il fait un vent à tout… ») ;
- elles sont souvent métaphoriques et idiomatiques, donc porteuses d'information culturelle ; leur sens n'étant pas toujours transparent, il faut les présenter en contexte et les expliquer comme tout autre item lexical ».

En plus de réduire le temps de recherche lexicale (le rappel de blocs de mots est plus rapide que la génération créative), l'usage d'unités polylexicales permet à l'apprenant de parler avec plus de fluidité, d'authenticité et de précision, et de parler comme un natif. Si l'apprenant est perçu comme étant un natif ou ayant une bonne maîtrise de la langue, son interlocuteur s'adressera à lui avec un langage plus riche, ce qui améliorera la qualité de l'input et donc de l'apprentissage (Théophanous, 2008). On a là encore affaire à l'Effet Matthieu.

La connotation et le registre d'un item lexical sont aussi à prendre en compte. Binon et Verlinde (2004 : 273) rappellent que privilégier l'acquisition du lexique ne veut pas dire « négliger les composantes stratégique, discursive, sociolinguistique et socioculturelle de la compétence de communication en FOS ». A ce titre il faut prendre en compte le registre d'un item lexical. Bogaards (1994) parle de *sens connotatif* ce qui n'est pas la même chose que la représentation sémantique de l'item lexical. La connotation est de l'ordre du style ou niveau de langue, tout en ajoutant de la valeur sémantique sans être grammaticale ou lexicale. Il ne suffit pas de connaître uniquement les sens dénotatifs des mots, car les sens dénotatif et connotatif sont complémentaires :

> « Loin de s'opposer, les sens dénotatif et connotatif, formant deux facettes d'une même réalité langagière, sont intimement liés. Selon le contexte, c'est tel aspect de l'un ou de l'autre qui est actualisé. Enfin, l'on peut dire qu'il n'y a guère d'[items lexicaux] neutres ou uniquement objecti[f]s ; en principe, chac[un] comporte, mais dans des proportions assez différentes, ses côtés dénotatifs et ses côtés connotatifs. Une véritable maîtrise du vocabulaire prend en compte cette réalité double » (Bogaards, 1994 : 30).

A ces sous-catégories liées à l'usage, nous ajouterons celle de la charge culturelle partagée (Galisson, 1991). En s'inscrivant dans le cadre de l'approche communicative, cet auteur plaide pour une prise en compte de la culture dans la didactique des langues. Quand on apprend une langue étrangère, on apprend aussi une culture étrangère, langue et culture allant de pair. L'apprentissage d'une culture passe obligatoirement par l'apprentissage du vocabulaire : « Le discours est chargé de culture à tous les niveaux (phonologique, morphologique, syntaxique, lexical — le gestuel coverbal en est lui-même tout imprégné), mais c'est encore le vocabulaire qui est le plus gros convoyeur de culture » (Galisson, 1991 :

159). La visioconférence semble être un contexte particulièrement propice à des échanges culturels et au développement de compétences culturelles (O'Dowd, 2000 ; Thorne, 2005).

La culture partagée, selon Galisson (1991), appartient à tous les membres d'une communauté[14] et leur permet de vivre ensemble. Elle est vécue et non décrite. Elle s'apprend en dehors de l'école et est difficilement appréhendée par les étrangers. Certains mots véhiculent cette charge culturelle partagée plus que d'autres. Galisson (1991 : 120) en donne la définition suivante :

> « J'appelle « Charge Culturelle Partagée » (C.C.P. !) la valeur ajoutée à leur signification ordinaire et pose que l'ensemble des mots à C.C.P. connus de tous les natifs, circonscrit la lexiculture partagée. Laquelle est toute désignée pour servir de rampe d'accès à la culture omniprésente dans la vie des autochtones et que les étrangers ont tant de mal à maîtriser — sans doute parce qu'elle n'est décrite, donc enseignée nulle part à ce jour : la culture partagée ».

Compte tenu de l'importance de la charge culturelle partagée et de sa difficulté à être appréhendée par les non-natifs, un enseignement explicite de celle-ci s'impose. Par exemple, dans le corpus que nous allons étudier, les apprentis-enseignants sont confrontés à l'item lexical « bobo » pour désigner un type de population et doivent faire comprendre ce concept qui porte une charge culturelle partagée importante à leurs apprenants irlandais (cf. chapitre 9).

Pour résumer, la connaissance d'un item lexical est pluridimensionnelle. De manière générale, connaître un item lexical implique que l'on connaisse sa forme, ses sens, et son usage. Dans notre corpus, les apprentis-enseignants ont tendance à se focaliser sur une ou deux facettes de la connaissance à la fois. Nous allons nous focaliser dans la section suivante sur les conditions qui favorisent l'acquisition de cette connaissance.

2.4 Qu'est-ce qui favorise l'apprentissage du vocabulaire ?

Jusqu'alors, nous avons décrit les dimensions de la connaissance d'un item lexical : forme, sens, et usage. Les procédés par lesquels les items lexicaux sont appris sont complexes, et maintenant nous allons nous attacher à les décrire afin d'être en mesure d'analyser les séquences d'explication lexicale. Toutefois, quelques mises en garde s'imposent. Bien que notre thèse porte sur les explications lexicales en interaction, il convient de rappeler que l'apprentissage délibéré du vocabulaire hors contexte par listes de mots semble être plus bénéfique pour la mémorisation à long terme et pour le développement de compétences productives que l'apprentissage incident (Folse, 2004). Par conséquent,

[14] Contrairement à la connotation, qui n'est pas collective selon Galisson (1991).

l'autonomie des apprenants est indispensable, car faute de temps, la plupart des mots doivent être appris en dehors du cours. Enfin, il ne faut pas perdre de vue que l'acquisition du vocabulaire n'est pas l'objectif premier des interactions pédagogiques que nous observons. L'apprentissage du vocabulaire, comme le rappelle Fasel Lauzon (2014 : 219) « doit plutôt être envisagé comme une sorte de 'valeur ajoutée' émergeant de certaines séquences d'explication ou d'activités qui les suivent, qui les rend particulièrement riches sur le plan des opportunités d'apprentissage ». Cependant, l'enseignement du vocabulaire est de première importance pour les raisons citées ci-dessus, et il y a certains facteurs qui sont susceptibles de privilégier l'apprentissage incident du vocabulaire en cours. C'est sur ceux-ci que nous allons nous focaliser dans cette section.

2.4.1 Focalisation sur la forme

Pour qu'il y ait l'apprentissage, on doit d'abord être conscient de ce que l'on perçoit (Schmidt, 1990). Selon cet auteur, qui est cité souvent dans des travaux sur l'acquisition du vocabulaire, le repérage (ou le « *noticing* ») est absolument essentiel, car l'apprentissage ne peut pas avoir lieu de manière inconsciente. Concrètement, cela veut dire que la forme d'un item lexical doit être repérée par l'apprenant avant d'être apprise (Groot, 2000). Lüdi (1991 : 199) considère « que la mémorisation de l'information lexicale dépend en partie de sa saillance dans l'interaction ». La forme d'un item lexical a plus de chances d'être repérée si elle est fréquente, saillante, inattendue, nécessaire pour la réalisation de la tâche en cours (Schmidt, 1990), ou mise en exergue par l'enseignant (Nation & Webb, 2011). Après avoir repéré la forme d'un item lexical, il faut focaliser de manière consciente sur celle-ci (Dobinson, 2001) car on apprend ce sur quoi on se focalise (Barcroft, 2006).

La représentation phonologique d'un item lexical est primordiale, même à la lecture (García, 2008 ; Marijanović, Billières, & Panissal, 2008 ; Nation, 2013 ; Schmitt, 2008; Tréville, 2000 ; Tréville & Duquette, 1996). Tréville (2000 : 63) affirme qu' « il existe une forte corrélation entre l'aptitude à construire des représentations phonologiques de mots nouveaux dans la mémoire et l'apprentissage du vocabulaire de la L2 ». Laufer (1994) explique que si deux mots se ressemblent phonologiquement[15], cela créera des interférences et rendra l'apprentissage plus difficile. Si les sons qui forment le mot en L2 n'existent pas dans la L1 de l'apprenant, le mot sera plus difficile à percevoir et à apprendre. Une bonne maîtrise du système phonologique de la L2 est utile pour l'acquisition du vocabulaire (Van der Linden,

[15] Par exemple, « magazine » et « magasin » (cf. chapitre 6, exemple 27 en page 210).

2006). Une bonne manière de maîtriser la prononciation d'un item lexical et par conséquent sa représentation phonologique est de le répéter à haute voix (Ellis & Beaton, 1993).

Si on reconnaît les racines et les affixes, cela peut aider la compréhension mais aussi induire l'apprenant en erreur. Tréville (2000) plaide pour l'enseignement du vocabulaire par l'analyse morphologique, qui permet à l'apprenant d'identifier quelques racines qu'il retrouvera par la suite dans d'autres mots. Il incombe donc aux enseignants de veiller à ce que les formes orales et écrites des items lexicaux soient soulignées et repérées. Nous analyserons dans le chapitre 6 la façon dont les apprentis-enseignants se focalisent avec une multiplicité de ressources multimodales sur la forme des items lexicaux et les fonctions pédagogiques liées à cette focalisation.

2.4.2 Focalisation sur le lien entre forme et sens

Après avoir repéré et appris la forme d'un item lexical, l'étape suivante est, naturellement, d'apprendre son sens.[16] Au niveau cognitif, l'encodage sémantique crée des traces plus durables que l'encodage phonologique (Baddeley, 1993) et implique un traitement cognitif plus profond (Bogaards, 1994). La focalisation sur le sens mérite notre attention car la façon dont le sens est enseigné a des répercussions à la fois sur la compréhension et sur la rétention à long terme.

2.4.2.1 Contextualisation

Si les mots sont rencontrés dans leur contexte, l'apprenant établit des liens entre les mots, les contextes et les tâches, ce qui renforce la mémorisation. Laufer & Shmueli (1997) donnent trois raisons pour lesquelles le contexte est important : (1) il est naturel de rencontrer des mots dans leur contexte ; (2) le contexte illustre les usages courants du mot ; (3) le contexte favorise le traitement profond (cf. *infra* section 2.4.3 en page 59) et offre une prise cognitive, ce qui renforce la mémorisation. Sur ce deuxième point, Tréville (2000 : 100) ajoute que le contexte donne des informations sur « les règles de distribution et les régularités de combinaison syntagmatique ». Le contexte est aussi important pour activer la bonne acception d'un mot polysémique. Un même mot peut avoir des sens différents selon le domaine qui est évoqué. Il faut donc activer un champ sémantique pour que le mot ait un sens (Tréville & Duquette, 1996). Néanmoins, si les mots ne sont pas nécessaires pour la compréhension du texte, ou pour la réalisation de la tâche, les mots risquent d'être ignorés et de ne pas être repérés. Pour être appris ils doivent d'abord être repérés et faire l'objet d'une

[16] Cela n'implique en rien que l'apprentissage dans le sens inverse, allant du sens vers la forme, soit impossible.

focalisation consciente. Laufer et Shmueli (1997) plaident donc pour une focalisation explicite et volontaire sur la forme des mots cibles rencontrés en contexte afin qu'ils soient appris.[17]

2.4.2.2 Groupement sémantique et thématique

La recherche a démontré que l'enseignement simultané de mots appartenant à un même champ sémantique est problématique, surtout en début d'apprentissage quand les mots en question sont peu connus (Erten & Tekin, 2008 ; Nation, 2000 ; Théophanous, 2008 ; Tinkham, 1997). Des mots sont liés sémantiquement s'ils sont hyponymes ou méronymes d'un même hyperonyme[18] : les couleurs, les jours de la semaine, ou les parties du corps. Les nouveaux mots sont souvent présentés en champs sémantiques par des manuels scolaires et par des enseignants de langues pour les raisons suivantes : (Erten & Tekin, 2008 ; Nation, 2000 ; Tinkham, 1997)

- Ce mode de présentation facilite le déroulement de tâches ;
- Le lexique mental en L1 et les savoirs semblent être organisés ainsi ;
- On comprend plus profondément le sens d'un mot si on voit les relations qu'il entretient avec les autres mots appartenant au même champ sémantique.

Par contraste, des mots sont liés thématiquement s'ils sont liés à un thème principal : *étang*, *sauter*, *nager*, *vert*, et *glissant* ont tous trait au thème *grenouille* (Tinkham, 1997). Alors que le regroupement de mots par thèmes facilite l'apprentissage, le regroupement de mots par blocs sémantiques semble rendre plus difficile cet apprentissage.

Tinkham (1997) a mené une expérience dans laquelle trois groupes d'étudiants ont dû apprendre des mots qui étaient liés sémantiquement, thématiquement, ou pas du tout, à l'oral et à l'écrit. Tinkham a trouvé que les groupes de mots qui étaient liés sémantiquement ont été appris avec plus de difficulté que les groupes de mots qui n'étaient pas liés du tout. D'autre part, les mots qui étaient liés thématiquement ont été appris avec plus de facilité que les mots qui ne l'étaient pas. Le regroupement sémantique serait nuisible pour l'apprentissage à l'oral aussi bien qu'à l'écrit, pour la reconnaissance autant que pour le rappel. L'auteur plaide donc pour que les mots-cibles soient présentés en groupes thématiques et non pas en groupes

[17] Cependant, l'apprentissage de mots hors contexte par listes est plus efficace et moins couteux en temps que l'apprentissage incident de mots en contexte selon Folse (2004).

[18] Un *hyponyme* est un mot dont le sens est plus spécifique que celui d'un autre. A l'inverse, un *hyperonyme* est un mot dont le sens est moins spécifique, plus général et comprend le sens d'autres mots plus spécifiques. Par exemple, « être humain » est un hyponyme de « mammifère », qui lui est un hyponyme d' « être vivant ». « Meuble » est un hyperonyme de « chaise ». Un *méronyme* est un terme qui entretien avec un autre terme une relation de partie à tout. Par exemple, « toit » est un méronyme de « maison ».

sémantiques. Erten & Tekin (2008) ont enseigné quatre groupes de vingt mots anglais à soixante enfants turcs. Deux groupes de mots étaient composés de mots sémantiquement liés. Les chercheurs ont trouvé que les mots qui n'avaient aucun lien sémantique entre eux ont été mieux mémorisés que les mots sémantiquement liés, et ce de manière significative, immédiatement après une séance d'apprentissage de quarante minutes et lors d'un post-test une semaine après l'apprentissage initial. Tout porte donc à croire que la présentation simultanée de mots sémantiquement liés rend l'apprentissage plus difficile. Les interférences se jouent non pas seulement au niveau phonologique mais aussi au niveau sémantique.

Bien que la présentation simultanée de nouveaux mots appartenant au même champ sémantique puisse induire l'apprenant en erreur, principalement en début d'apprentissage ou quand l'un des mots est peu connu, l'activation d'un champ sémantique peut être utile :

> « d'un point de vue linguistique, pour cerner le sens d'une unité, pour délimiter les zones de recoupement avec des mots voisins et pour saisir les règles d'emploi de cette unité ; d'un point de vue psychologique, pour favoriser la compréhension et la rétention de mots à l'intérieur de regroupements organisés qui correspondent à un système logique de représentation des connaissances et qui coïncident, on l'espère, avec la structure de notre lexique mental » (Tréville & Duquette, 1996 : 27).

C'est pour ces raisons, entre autres, que Nation (2000) mentionne l'utilité de voir des mots en champs sémantiques lorsque les connaissances lexicales sont plus avancées. Folse (2004) estime aussi qu'il est judicieux d'utiliser des tâches qui juxtaposent des mots appartenant au même champ sémantique pour la révision, mais non pour l'enseignement initial. Théophanous (2001) suggère que l'on attende qu'un vrai besoin communicatif se fasse sentir avant d'enseigner en même temps des synonymes ou des mots qui sont formellement ou sémantiquement proches. Nation (2000) propose deux moyens de pallier les effets d'interférence : espacer de quelques jours l'enseignement de mots sémantiquement proches, et utiliser des contextes très différents pour l'enseignement de chaque mot. Si les effets d'interférence se manifestent, on peut faire prendre conscience aux apprenants de ce phénomène, et encourager l'usage de techniques mnémoniques comme celle du mot-clé (Nation, 2000) (cf. *infra* section 2.4.6 en page 70).

2.4.2.3 Définitions simples

Étant donné la nature pluridimensionnelle d'un item lexical, il est impossible de tout apprendre en une seule rencontre car les apprenants ne peuvent traiter que de petites quantités d'information à la fois (Nation, 2013). En effet, les êtres humains disposent d'une capacité cognitive limitée dans le sens où parmi tout ce qui est présenté, on ne peut traiter et retenir que de petits morceaux (Mayer, 2001). Donner trop d'informations à la fois entraîne une surcharge cognitive et nuit à l'apprentissage (Sweller, 2005a). Par conséquent, les

explications courtes et simples qui ne répondent qu'à ce qui est demandé par l'apprenant sont à privilégier. Les définitions simples respectent le *principe de redondance* (Mayer, 2001 ; Sweller, 2005b) selon lequel donner plus d'informations aux apprenants nuit à l'apprentissage. Il y a *redondance* si on présente les mêmes informations avec deux modes différents, par exemple textuellement et par images (cf. chapitre 3), ou si on présente des informations supplémentaires par un même mode, par exemple oralement. Si les informations supplémentaires sont redondantes et ne sont pas nécessaires à la compréhension du message, la suppression de celles-ci peut faciliter l'apprentissage.

Chaudron (1982), qui a mené une étude dans quatre lycées au Canada, a trouvé qu'un grand problème pour la compréhension était le décodage d'explications trop complexes ou trop redondantes. Les définitions longues et compliquées embrouillaient l'apprenant car celui-ci ne savait pas si l'explication était redondante ou s'il existait de nouvelles informations. Ellis (1995), qui a analysé les effets sur la compréhension de l'input pré-modifié et de l'input modifié dans l'interaction, a trouvé que l'input modifié est efficace *à condition que les définitions soient courtes et simples*. Les explications longues, complexes et trop élaborées entraînent une surcharge cognitive. Quand un apprenant reçoit une explication courte et simple, celle-ci ne devrait pas être critiquée comme étant insuffisante, mais doit être reconnue comme étant une étape importante vers un apprentissage complet de l'item lexical (Nation, 2013).

2.4.2.4 Gloses

Une bonne manière d'assurer la compréhension de mots-cibles ou de mots potentiellement problématiques est de les accompagner de courtes définitions lorsqu'ils apparaissent dans un discours oral ou dans un texte. Par exemple, « je suis *restauratrice*, donc quelqu'un qui fait la cuisine »[19]. On trouve souvent des gloses dans les marges des pages des manuels scolaires et dans le discours oral de l'enseignant de langue. L'utilité de l'usage de gloses a été étudiée expérimentalement. Hulstijn *et al.* (1996) ont mené une expérience où ils ont donné à soixante dix-huit étudiants néerlandais avancés apprenant le français un court texte de Guy de Maupassant dans lequel se trouvaient seize mots cibles inconnus de tous les participants. Les chercheurs ont divisé les sujets en trois groupes : un groupe contrôle, où il n'y avait aucune explication des mots cibles, un groupe « gloses », où les mots cibles ont été définis en français ou en néerlandais dans les marges de la page, et un groupe « dictionnaire »

[19] Mélissa a énoncé cette phrase lors de la cinquième semaine des interactions (cf. chapitre 7, exemple 9 en page 228).

où les sujets avaient droit à l'usage d'un dictionnaire pendant la lecture. Ils ont trouvé que le groupe « gloses » a eu des résultats significativement plus élevés que les deux autres groupes pour l'apprentissage des mots cibles. Il n'y avait aucune différence significative entre les groupes « contrôle » et « dictionnaire ». L'une des raisons pour ce résultat est que les sujets qui disposaient d'un dictionnaire ne l'ont consulté que pour 12-15% des mots-cibles. Cependant, les mots qui ont été consultés ont été mieux appris que les mots qui faisaient l'objet de gloses. Les chercheurs ont trouvé aussi une influence positive du nombre de répétitions dans le texte sur l'acquisition des mots, ainsi qu'un effet positif associé à l'usage de traductions. Néanmoins, plus de la moitié de tous les mots appris ont été très vite oubliés après l'expérience.

Les gloses favorisent l'acquisition du vocabulaire aussi bien à l'oral qu'à l'écrit. Brett, Rothlein, et Hurley (1996) ont étudié les effets des gloses à l'oral sur l'acquisition du vocabulaire chez les enfants. Ils ont trouvé que les élèves qui recevaient de courtes définitions à l'oral des mots cibles rencontrés en écoutant une histoire ont appris significativement plus de mots qu'un groupe contrôle qui n'avait pas écouté l'histoire et qu'un groupe qui écoutait l'histoire mais à qui aucune définition n'avait été donnée. Cette différence était toujours vraie six semaines après l'écoute. Tout porte à croire que les gloses, qu'elles soient à l'écrit ou à l'oral, en langue étrangère ou en langue maternelle, favorisent l'acquisition du vocabulaire.

2.4.2.5 Traductions

Nation (2013) estime que la traduction est le type d'explication le plus simple, le plus concis et le plus familier. Celle-ci peut résoudre rapidement et efficacement un problème de compréhension. Cet auteur cite Hummel (2010) qui a trouvé que la traduction renforce les liens d'un mot nouvellement appris et Altarriba & Knickerbocker (2011) qui ont trouvé que l'apprentissage par traductions réduit le temps de réponse par rapport à l'usage de photos. Laufer & Shmueli (1997) ont trouvé que la traduction est la méthode la plus efficace pour apprendre de nouveaux mots en lisant et favorise la mémorisation. Pour ceux-ci, la traduction de mots hors contexte a même été plus efficace que la définition de mots en L2 en contexte. Tréville (2000 : 68) résume en disant que

> « Le recours à la traduction pour enseigner le vocabulaire (par paires de mots), loin d'être banni, s'est même révélé, à la suite de recherches empiriques, comme la manière la plus efficace de présenter le sens des mots et la plus favorable à la rétention (Nation, 1982, Carter, 1987, McCarthy, 1990) ».

Laufer & Shmueli (1997) donnent quatre arguments en faveur de la traduction :

- Les traductions sont souvent très précises ;

- L'expérience humaine est en quelque sorte universelle, ce qui fait que les mêmes concepts sont lexicalisés de manière similaire dans différentes langues ;
- Dans le cas où un concept serait lexicalisé différemment par différentes langues, par exemple « to know » en anglais et « connaître » et « savoir » en français, la traduction est nécessaire afin que l'apprenant puisse se rendre compte que le concept renvoie à un mot dans une langue et à deux mots dans l'autre ;
- La plupart des apprenants utilisent des dictionnaires bilingues.

La traduction comme stratégie d'explication n'est pas sans critiques. Par exemple, « si l'appariement initial d'un mot à un concept passe par la traduction, les liens entre la forme du mot et le sens risquent d'être dépendants de la L1 […], et le chemin d'accès trop long pour une utilisation efficace du mot en situation de communication » (Hilton, 2002 : 11)[20]. Folse (2004) liste deux critiques : les traductions exactes sont parfois inexistantes, et la traduction ne peut pas donner une compétence active immédiate. Pour réfuter ces deux critiques, Folse souligne que le nombre de mots sans traductions exactes est petit, et qu'aucune technique d'enseignement n'est capable de créer une compétence active immédiate. Il plaide pour que la traduction soit encouragée et privilégiée. Trois critiques sont soulevées par Nation (2013) : la traduction est indirecte, réduit le temps passé sur la L2, et comme le note Folse (2004), encourage l'idée qu'il existe une équivalence entre les mots en L2 et leurs traductions en L1. La réponse de Nation est la suivante : bien que vraies, ces critiques sont aussi valables pour les autres stratégies d'explication. Il n'existe pas une correspondance exacte et absolue entre un mot en L2 et sa définition en L2, ni entre un mot en L2 et un objet censé le représenter. L'usage d'images et de démonstrations réduisent le temps passé sur la L2 au même titre que la traduction. L'avantage de la traduction est qu'elle est rapide et facilement comprise par l'apprenant.

Kelly & Bruen (2015) décrivent la controverse qu'il existe autour de la traduction. La traduction a une mauvaise réputation en didactiques des langues car cette stratégie d'explication est souvent associée à des méthodes de grammaire-traduction. Avec l'émergence de l'approche naturelle, les méthodes favorisant la focalisation sur le sens plutôt que sur la forme, et l'approche communicative, la traduction est tombée en désuétude. Ces auteurs ont récemment étudié des attitudes à l'égard de la traduction auprès de douze enseignants de langue, six japonais et sept allemands. Les interviews ont révélé que les attitudes étaient largement positives, les enseignants estimant que la traduction facilitaient

[20] La pagination correspond au manuscrit auteur.

l'apprentissage du vocabulaire, la compréhension du contexte et de la langue en générale, l'acquisition de structures grammaticales, le repérage de lacunes langagières, une conscientisation d'enjeux culturels, l'adaptation aux différents styles d'apprentissage, et permettaient aux apprenants de prendre du plaisir à apprendre. Il se peut donc que les attitudes à l'égard de la traduction soient en train d'évoluer.

2.4.2.6 Rappel

Chaque association réceptive ou productive entre forme et sens renforce ce lien, et la mémorisation à long terme est davantage favorisée si les rappels sont espacés (Baddeley, 1993). Hulstijn *et al.* (1996) ont trouvé aussi une corrélation positive entre le nombre de répétitions d'un mot-clé dans un texte et la mémorisation. Nation & Webb (2011) font une distinction entre rappel productif et rappel réceptif, et soulignent que le rappel productif est plus difficile que le rappel réceptif. Ils mentionnent qu'une façon d'encourager le rappel chez l'apprenant est de ne pas donner la forme ou le sens tout de suite. Selon Folse (2004), les rappels provoqués par l'enseignant favorisent l'acquisition du vocabulaire.

Pour résumer, les explications efficaces sont accompagnées d'une focalisation sur la forme de l'item lexical. Le lien entre forme et sens est établi. Les mots sont rencontrés en contexte et ne sont pas présentés avec des synonymes, antonymes, ou d'autres mots qui sont sémantiquement liés. Les collocations, registres et connotations sont abordés. Les explications sont courtes, concises, simples, et ne sont pas trop complexes, élaborées ou redondantes. Les traductions sont à favoriser. A la suite de l'explication initiale, les rappels productifs et réceptifs sont à privilégier. Enfin, les apprenants devraient être encouragés à noter par écrit tous les nouveaux mots qu'ils rencontrent. Ils doivent être aussi encouragés à interroger l'enseignant lorsqu'ils rencontrent une panne lexicale productive ou réceptive.

La plupart des études que nous avons citées dans cette section sont des études expérimentales avec des variables contrôlés : le nombre d'occurrences dans un texte, la présence ou non de gloses, la traduction versus une définition en langue cible, etc. On peut se demander dans quelle mesure ces résultats sont transférables dans une situation écologique comme la nôtre. C'est une des raisons pour lesquelles nos analyses se veulent descriptives. L'autre raison est que nous n'avons pas administré à nos sujets un post-test, ce qui nous aurait permis de mesurer l'efficacité des stratégies que nous allons décrire dans nos analyses. Nous donnerons en conclusion des propositions pour d'éventuels études expérimentales en visioconférence.

Cette section s'est focalisée sur les caractéristiques des explications lexicales susceptibles de favoriser l'acquisition. Nous allons nous concentrer maintenant sur d'autres facteurs propices à l'apprentissage, notamment le traitement profond et la négociation du sens. Ces phénomènes nous serviront de bases théoriques pour le chapitre 8 qui analysera l'implication des apprenants dans les séquences d'explication lexicale.

2.4.3 Traitement profond

Craik & Lockhart (1972) ont inventé un cadre théorique pour comprendre les processus mentaux qui créent des traces mémorielles. La profondeur du traitement fait référence au niveau d'analyse sémantique ou cognitive. Ces auteurs postulent que la durée de mémorisation d'un item lexical est liée à la profondeur du traitement et non pas à la vitesse du traitement ou à d'autres critères. Le niveau de traitement peut varier, allant d'un traitement peu profond d'un mot qui consisterait à regarder simplement sa typographie, à un traitement moyen qui consisterait à trouver des mots avec lesquels il rime, jusqu'à un traitement profond qui consisterait à se demander si ce mot peut faire partie d'une catégorie ou s'insérer dans une phrase donnée. Les auteurs postulent que c'est le niveau d'analyse, ou la profondeur de traitement, qui exerce une influence sur la mémorisation d'un mot. Plus la mémorisation est profonde, meilleur sera le rappel ultérieur lors d'une tâche réceptive ou productive.

Ce cadre théorique a été vérifié à l'aide de dix études expérimentales effectuées par Craik & Tulving (1975). En faisant une synthèse de ces dix expériences, les auteurs concluent que ce qui détermine le niveau de rappel ou de reconnaissance d'un mot n'est ni l'intention d'apprendre, ni l'effort exercé, ni la difficulté de la tâche, ni le temps de réflexion sur les items lexicaux, ni le temps d'entraînement passé sur les items, mais plutôt la nature qualitative de la tâche et le type d'opérations et manipulations réalisées sur les items lexicaux. On ne peut pas calculer la performance mnémonique comme une fonction du nombre d'attributs, car c'est la nature qualitative de ces attributs qui seule compte. Baddeley (1993 : 180) critique cette approche en disant qu'il est difficile de mesurer les niveaux de traitement sans succomber à des raisonnements circulaires : « si une manipulation A conduit à un bon apprentissage, alors elle est supposée requérir un traitement profond, si l'apprentissage est faible, alors le traitement a dû être superficiel ». Le cadre théorique proposé par Craik et Lockhart nous semble être un bon point de départ pour mesurer l'implication cognitive des apprenants.

Quels sont donc les facteurs contribuant à un traitement profond d'un item lexical ? Les sous-sections suivantes répondent à cette question.

2.4.3.1 Output provoqué par l'enseignant

Swain (1995) souligne l'importance de l'output et s'inscrit en faux contre l'idée de Krashen (1985) selon laquelle l'output n'est pas plus qu'un indicateur de l'acquisition. Swain (1985, 1995) argumente en faveur d'une conception acquisitionniste de l'output. Selon elle, l'output facilite l'apprentissage du lexique et de la langue en général :

- La production langagière améliore l'aisance ;
- L'output encourage le repérage de formes langagières ;
- En produisant du langage, on le traite plus profondément ;
- On repère des lacunes langagières qui doivent être comblées. On peut feindre la compréhension mais non pas la production. On se rend compte des écarts entre ce qu'on *veut* dire et ce qu'on *peut* dire. Cette tension entre l'interlangue de l'apprenant et la compétence linguistique native souhaitée est nécessaire pour l'acquisition (Porquier & Py, 2008) ;
- L'output permet à l'apprenant de tester des hypothèses par rapport à la compréhension et l'exactitude d'un énoncé. Ces hypothèses seront par la suite validées ou corrigées par l'enseignant ;
- L'output possède une fonction métalinguistique qui permet aux apprenants de contrôler et internaliser des savoirs linguistiques, surtout quand ceux-ci réfléchissent sur la langue en effectuant une tâche nécessitant une compréhension du sens de la langue.

Les apprenants ont donc intérêt à provoquer l'output chez l'apprenant afin qu'il se rende compte de lacunes lexicales et réutilise des items lexicaux une fois appris.

2.4.3.2 Porter un jugement sur un item lexical

Une façon de traiter un mot de manière profonde est de porter un jugement sur celui-ci. Selon Baddeley (1993), les mots qui sont jugés sur deux dimensions (par exemple agréable/désagréable, fort/faible) sont mieux mémorisés que les mots qui ne sont jugés que sur une seule dimension. Cet effet est amplifié quand il y a une grande différence entre les dimensions de jugement. Calaque (2000 : 23) plaide pour une « prise en compte de l'expérience subjective et du rapport à la langue ». Bourdon-Larger (2008) pense aussi que le

fait de porter un jugement sur un mot et de mener une réflexion sur sa valeur affective facilite la mémorisation.

2.4.3.3 Usage génératif

Joe (1995, 1998) s'est attachée à explorer la théorie de la profondeur de traitement de Craik & Tulving (1975) en mesurant l'effet de l'usage génératif sur l'acquisition du vocabulaire. L'usage génératif consiste à utiliser un mot d'une façon autre que la façon dont il a été appris. C'est par exemple le fait d'utiliser un mot dans un contexte autre que le contexte dans lequel il a été appris, ou le fait d'utiliser un sens du mot qui est autre que le sens initialement appris. C'est d'ailleurs une des raisons pour lesquelles on parle de *mots* ou d'*items lexicaux* plutôt que d'*unités lexicales* (Bogaards, 1994). En observant quarante-huit adultes apprenant l'anglais en Nouvelle Zélande, Joe (1998) a trouvé une corrélation significative et positive entre usage génératif et acquisition, surtout pour les mots qui étaient inconnus au début de l'expérience. Ce qu'il faut retirer de cette expérience est que plus il y a du traitement génératif, plus il y a d'apprentissage du vocabulaire. L'usage génératif est facilité notamment par la tâche. Ce constat corrobore celui de Baddeley (1993) : que la réalisation de tâches qui requièrent le rappel de mots peut favoriser l'acquisition incidente du vocabulaire, même si on ne se focalise pas explicitement sur le vocabulaire. L'usage génératif est souvent difficile : comme le rappelle Leeman (2005), parfois on peut deviner le sens d'une autre acception d'un mot, et parfois non. Fasel Lauzon (2014) préconise aussi la mise en place d'activités qui favorisent la réutilisation du terme nouvellement appris dans des contextes syntaxico-sémantiques variés.

2.4.3.4 Charge d'implication

Laufer & Hulstijn (2001), en se basant sur les travaux de Craik & Lockhart (1972) ont élaboré une hypothèse très connue qui s'appelle l'*involvement load hypothesis* (hypothèse de la charge d'implication). Selon ces auteurs, le niveau d'implication de l'apprenant, et donc l'apprentissage incident du vocabulaire, serait fonction de trois facteurs : besoin, recherche et évaluation, ces deux derniers étant cognitifs. Pour chacun des facteurs, il y a trois niveaux : absent, modéré et fort. Les auteurs supposent que plus il y a la présence de ces trois facteurs, plus la charge d'implication est forte, et mieux les mots seront appris et retenus. Ils se demandent si la qualité du traitement des mots est plus importante que la quantité (nombre de répétitions ou de rencontres). Ces facteurs seront explorés dans le chapitre 8.

Le besoin : celui-ci constitue la dimension non-cognitive de l'hypothèse. Il y a besoin si la compréhension du mot est nécessaire pour la réalisation de la tâche, ou si l'apprenant en

a besoin pour s'exprimer. Le besoin est modéré s'il vient d'un facteur externe : l'enseignant demande à l'apprenant de s'exprimer d'une certaine manière. Le besoin est fort si l'apprenant se l'impose : l'apprenant veut s'exprimer mais ne le peut pas faute de connaissances lexicales.

La recherche et l'évaluation : celles-ci constituent les dimensions cognitives de l'hypothèse et surviennent quand on décide consciemment de porter attention sur la relation forme-sens d'un item lexical. Il y a recherche quand on tente de trouver le sens d'un mot inconnu en L2 ou quand on tente de trouver la forme d'un mot à partir de son sens. Ces deux phénomènes sont aussi appelés respectivement rappel réceptif et rappel productif. La recherche est modérée s'il s'agit d'un rappel réceptif, et fort s'il s'agit d'un rappel productif.

Il y a évaluation quand on compare un mot donné à d'autres mots, ou quand on compare les différentes acceptions d'un mot entre elles. Par exemple, quand on consulte un dictionnaire et qu'on trouve plusieurs sens pour un seul mot, on doit sélectionner le sens le plus adapté au contexte donné. Autre exemple : l'apprenant consulte un dictionnaire bilingue et trouve 3 mots en L2 pour un seul mot en L1 recherché. Il doit alors comparer les trois mots entre eux pour sélectionner le mot qui est syntaxiquement et/ou sémantiquement approprié pour la collocation souhaitée. L'évaluation implique donc une décision sélective au niveau de la sémantique ou de la syntaxe. L'évaluation est modérée si l'apprenant doit comparer les différents sens d'un mot entre eux, et est forte si l'apprenant doit comparer des mots entre eux afin d'en sélectionner les meilleurs lors de la formulation d'une phrase originale. Cette méthode d'analyse a été louée par nombre de chercheurs car c'est une manière simple et quantifiable d'analyser les techniques d'enseignement du vocabulaire. Nous discuterons dans le chapitre 8 de la façon dont ces facteurs d'implication peuvent être favorisés par les apprentis-enseignants.

Nation & Webb (2011) résument la technique d'évaluation de (Laufer & Hulstijn, 2001) ainsi :

Besoin	Aucun besoin	0
	Extrinsèque	1
	Intrinsèque	2
Recherche	Aucune recherche	0
	Réceptive (on cherche le sens)	1
	Productive (on cherche la forme)	2
Évaluation	Aucune évaluation	0
	Décontextualisée	1
	Contextualisée	2

Figure 3 : Grille d'analyse de Laufer & Hulstijn (2001) synthétisée par Nation & Webb (2011)

Les chiffres à droite du tableau représentent le nombre de points accordés à chaque critère (besoin, recherche, et évaluation). Cette grille de notation est utilisée pour comptabiliser l'efficacité de différentes tâches afin de comparer leur utilité pour l'apprentissage du vocabulaire. Par exemple, un besoin extrinsèque vaut un point car c'est un besoin modéré, tandis qu'un besoin intrinsèque vaut deux points car c'est un besoin fort. Nous discuterons dans le chapitre 8 de la façon dont ces critères d'implication peuvent être favorisés par les apprentis-enseignants.

2.4.3.5 Participation active

La discussion et la participation active aident l'apprentissage du vocabulaire, car quand on discute un mot, on le traite plus profondément que quand on le reconnait comme la bonne réponse sur une liste (Stahl & Vancil, 1986). Quand c'est l'apprenant qui demande une explication, le *besoin* est plus fort, ce qui favorise l'acquisition (Laufer & Hulstijn, 2001). Brouwer (2003) a trouvé que l'acquisition est favorisée quand les participants sont invités à participer activement dans des recherches lexicales. L'acquisition du vocabulaire est aussi favorisée quand ce sont les apprenants qui expliquent. Newton (2001) note qu'il est avantageux de laisser les apprenants travailler entre eux car l'étendue lexicale d'un groupe est toujours plus large et riche que l'étendue lexicale d'un individu. Folse (2004) mentionne aussi l'utilité de laisser l'apprenant le plus fort aider l'apprenant le plus faible. Selon Fasel Lauzon (2014 : 203), l'apprentissage est favorisé lorsque les apprenants sont impliqués dans le noyau de l'explication (cf. chapitre suivant) :

> « Le lien entre explication et apprentissage est double : on peut apprendre en expliquant et apprendre à expliquer […]. En conséquence, la participation des élèves aux séquences d'explication en classe peut les amener à développer des connaissances relativement à des contenus (ceux-là mêmes qui sont expliqués dans la séquence) mais peut également leur permettre de développer leur compétence d'explication (qui relève de la compétence d'interaction […]) »

Fasel Lauzon (2014 : 273) réaffirme ainsi l'importance de la participation active des apprenants aux séquences d'explication lexicale :

> « Concernant les explications de vocabulaire, les séquences qui apparaissent les plus propices non seulement pour la résolution du problème de compréhension mais également pour la mémorisation du mot expliqué et de ses conditions d'emploi sont celles dans lesquelles les élèves sont encouragés a) à participer activement à la construction du noyau de l'explication, et/ou b) à utiliser de manière récurrente le terme ayant fait l'objet de l'explication ».

Dobinson (2001) a aussi trouvé que l'interaction facilite l'apprentissage du vocabulaire, mais elle a trouvé que l'apprentissage incident en classe de langue est très aléatoire et idiosyncratique. Certains apprenants dans son étude de vingt-quatre apprenants d'anglais âgés de seize à quarante ans ont appris très peu de vocabulaire tandis que d'autres

ayant suivi la même formation en ont appris beaucoup. Certains mots ont été beaucoup négociés mais n'ont pas été appris, tandis que d'autres mots n'ayant été mentionnés qu'une seule fois ont été largement appris.[21]

Malgré l'importance de la participation active, celle-ci n'est pas toujours nécessaire car l'apprentissage de mots semble être le résultat de processus mentaux qui ne sont pas toujours visibles et qui existent même sans participation active. Avant de participer, l'apprenant doit formuler une réponse silencieusement dans sa tête, et cette activité mentale a lieu même si l'apprenant n'a pas l'occasion de prendre la parole et participer activement. En remarquant ce phénomène, Stahl & Clark (1987) se sont demandé si la participation active était vraiment nécessaire pour l'apprentissage du vocabulaire ou si l'anticipation seule suffisait. Pour répondre à cette question, ils ont divisé un groupe d'élèves à l'école primaire en trois groupes, plus un groupe contrôle sans apprentissage :

1. Un groupe « écoute » : l'enseignant leur a dit qu'ils n'allaient pas participer et qu'ils étaient là pour apprendre en écoutant. L'enseignant ne leur posait pas de questions.
2. Un groupe « interrogé » : l'enseignant leur a dit qu'ils allaient participer et l'enseignant leur posait des questions pour qu'ils participent.
3. Un groupe « ignoré » : l'enseignant leur a dit qu'ils allaient participer mais l'enseignant ne les a pas interrogés.

Pendant trois jours, ces groupes apprenaient du vocabulaire issu d'un texte de 600 mots dont le thème était l'astronomie. Les élèves changeaient de groupe chaque jour. Les résultats du post-test de l'étude ont montré que les groupes « interrogé » et « ignoré » ont eu de meilleurs résultats que le groupe « écoute », mais qu'il n'y avait aucune différence significative entre les groupes « interrogé » et « ignoré ». Par conséquent, la participation n'est pas si nécessaire que cela ; l'anticipation d'être interrogé par l'enseignant suffit pour qu'il y ait apprentissage du vocabulaire. Ces auteurs proposent donc que l'enseignant pose une question avant de nommer un apprenant pour répondre afin que tous les apprenants formulent des réponses. Les études de Newton (1995) et de Dobinson (2001) corroborent ce résultat que la participation n'est pas indispensable pour l'apprentissage du vocabulaire.

Ce que nous retenons des études susmentionnées c'est que la participation active aide l'apprentissage, mais que la participation active par chaque apprenant n'est pas indispensable. Toutefois, il doit y avoir négociation et participation pour que les apprenants qui ne

[21] L'inconvénient de ce type d'études est que l'on ignore quels mots ont été appris en dehors des cours.

participent pas puissent observer l'interaction et anticiper une participation éventuelle. Il incombe donc aux enseignants de veiller à ce qu'au moins un des apprenants participe aux explications lexicales. Dans notre corpus, les apprentis-enseignants n'ont qu'un ou deux apprenants chacun, et doivent faire en sorte que ces deniers participent à des séquences d'explication lexicale. Nous verrons dans les chapitres d'analyse que les apprenants sont très sollicités par les apprentis-enseignants, et parfois par leurs binômes.

2.4.4 Négociation du sens

Dans l'interaction, « chaque élément de la conversation peut être objet de négociation, le sens d'un mot aussi bien que le type d'activité à accomplir ou l'interprétation d'une activité accomplie » (Gülich, 1991 : 335). Dans le cadre de notre thèse, nous nous intéressons particulièrement à la négociation d'items lexicaux, car ce sont ceux-ci qui déclenchent le plus grand nombre de séquences de négociation du sens (de Pietro, Matthey, & Py, 1989 ; Ellis, Tanaka, & Yamazaki, 1994 ; Pica, 1994 ; Smith, 2003).

La négociation du sens est propice à l'acquisition du vocabulaire car elle permet l'émergence de bien des facteurs listés ci-dessus, notamment le repérage de lacunes lexicales, la focalisation sur la forme, la focalisation sur le sens, la modification de l'input, et la participation active (Long, 1981, 1983). Hatch (1978) fut l'une des premières à s'intéresser à l'interaction comme étant la cause et non pas le résultat de l'apprentissage langagier. Ce renversement des choses à donné lieu à quantité d'études sur la négociation du sens et son rôle dans l'acquisition langagière. Pica (1994 : 494) définit la négociation comme :

> « la modification et restructuration de l'interaction qui a lieu quand les apprenants et leurs interlocuteurs anticipent, perçoivent ou expérimentent des difficultés au niveau de la compréhensibilité du message. En négociant, ils réalisent un travail linguistique afin d'atteindre la compréhensibilité dont ils ont besoin par répétition exacte, ajustement de la syntaxe, changement des mots, ou modification de la forme et du sens par toute autre nature que ce soit » (notre traduction)[22].

Selon cette auteure, l'interaction est avantageuse pour l'apprentissage des langues car elle est la garante de trois choses qui sont nécessaires pour l'acquisition langagière :

1. L'input compréhensible ;
2. L'output modifié par l'apprenant ;
3. La focalisation sur la forme.

[22] « the modification and restructuring of interaction that occurs when learners and their interlocutors anticipate, perceive, or experience difficulties in message comprehensibility. As they negotiate, they work linguistically to achieve the needed comprehensibility, whether repeating a message verbatim, adjusting its syntax, changing its words, or modifying its form and meaning in a host of other ways » (Pica, 1994 : 494).

Comment sait-on que c'est bien la négociation qui garantie l'acquisition, et non pas seulement l'input compréhensible ? Le débat autour de l'input pré-modifié et l'input modifié dans l'interaction par la négociation a fait couler beaucoup d'encre. Ces deux types de modifications interactionnelles ont pour objectif de rendre l'input compréhensible et ont des caractéristiques en commun. Pica et al. (1987) distinguent deux types d'input modifié ou simplifié. D'une part, il y a ce qu'ils appellent l'input pré-modifié, c'est-à-dire l'input qui est rendu plus simple par des répétitions, paraphrases, vocabulaire réduit, connecteurs saillants, phrases courtes, complexité réduite, absence de clauses subordonnées, etc., avant qu'il ne soit présenté à l'apprenant. Ces auteurs citent plusieurs études qui ont démontré que la compréhension est améliorée quand l'input est pré-modifié. D'autre part, il y a ce que Pica *et al.* appellent l'input qui est modifié durant l'interaction par la négociation. Dans ce cas, les interlocuteurs restructurent leurs discours afin d'arriver à une compréhension mutuelle. Ces modifications sont initiées par ce qu'ils appellent des opérations inter-factionnelles, c'est-à-dire des demandes de clarification, vérifications de compréhension, et demandes d'aide. En postulant que la compréhension sera plus élevée dans cette seconde condition (l'input est modifié durant l'interaction par la négociation), Pica *et al.* (1987) disent avoir été les premiers à tester empiriquement cette hypothèse.

Ces chercheurs ont administré à seize apprenants d'anglais du niveau intermédiaire une tâche où les sujets devaient écouter les consignes d'un natif afin de choisir et placer quinze items (plantes, animaux et humains) sur un tableau illustré par une scène de nature. Huit sujets ont été placés dans un premier groupe où les consignes ont été pré-modifiées (simplifiées par des moyens listés ci-dessus) mais les sujets n'avaient pas le droit de poser des questions. Les huit sujets du second groupe ont reçu des consignes non-modifiées, c'est-à-dire destinées à des natifs, mais les sujets ont été encouragés à demander de l'aide et à interagir avec le natif. Les chercheurs ont trouvé que l'interaction et la négociation dans le second groupe ont produit plus d'input avec plus de redondance par rapport à l'input pré-modifié du premier groupe. Contrairement à une des hypothèses formulées par les chercheurs, l'input issu de la négociation était plus complexe que l'input pré-modifié. Le résultat le plus important est que les scores (sélection et emplacement des objets réussis) ont été significativement plus élevés pour le second groupe (groupe négociation). C'est-à-dire que l'input modifié dans l'interaction par la négociation du sens est plus efficace que l'input pré-modifié, bien que les deux aient des caractéristiques similaires.

Après avoir démontré empiriquement que la redondance et la quantité de l'input aident l'acquisition, et que ces modifications sont réalisées par la négociation, les chercheurs plaident pour que les enseignants demandent à leurs apprenants s'ils ont compris, et pour qu'ils les encouragent à initier des demandes de clarification et de confirmation. Ils argumentent en faveur d'une réorientation des rôles des participants dans les cours de langues de façon à ce que les apprenants puissent assumer leur part de responsabilité dans l'apprentissage, ce qui encourage l'interaction et améliore ainsi la compréhension de l'input. C'est une des raisons pour lesquelles la symétrie des participants en visioconférence est si importante.

Suite au constat que la plupart des séquences de négociation du sens sont déclenchées par un item lexical, Ellis *et al.* (1994) se sont demandé d'une part si l'input modifié par la négociation amènerait à une meilleure compréhension de mots en L2 que l'input non-modifié ou pré-modifié, et d'autre part si l'input modifié par la négociation amènerait à une meilleure rétention de mots. Pour répondre à ces questions, ils ont mis en place deux études de 79 et de 129 étudiants qui apprenaient l'anglais à Saitama et à Tokyo. Un pré-test et deux post-tests ont évalué leur connaissance de dix-neuf mots inconnus. Pour la tâche, les apprenants ont dû écouter des consignes pour placer un objet de cuisine sur une image matrice. Il y avait trois sous-groupes : le groupe A, qui écoutait des consignes non-modifiées, le groupe B, qui écoutait de l'input modifié tiré d'une discussion préalable avec d'autres sujets dans laquelle il y avait de la négociation, et le groupe C, où les apprenants pouvaient poser des questions et négocier l'input. En plus de calculer la redondance et la complexité de l'input des trois groupes, ils ont calculé la compréhension et l'acquisition des mots cibles via deux post-tests où les sujets devaient traduire des mots et nommer des images. Pour les deux groupes, l'input négocié dans l'interaction a amené à de meilleurs scores de compréhension que l'input pré-modifié ou non-modifié et ce de manière significative. Néanmoins, ces scores élevés de compréhension ne se sont pas traduits par des meilleurs scores de rétention lors du test de suivi. Cela veut dire que la compréhension ne garantit pas l'acquisition, car les processus de compréhension et d'apprentissage diffèrent. Un autre résultat intéressant de cette étude est que ceux qui ont participé activement n'ont pas appris mieux que ceux qui écoutaient sans participer, ce qui corrobore le constat fait par Stahl & Clark (1987) et Stahl & Vancil (1986).

Parfois la négociation n'est pas la source principale de l'acquisition. Newton (2001) explique que la négociation n'est pas toujours favorable à l'apprentissage si les apprenants ne se focalisent que sur la forme, ce qui est souvent le cas lors des tâches d'informations

partagées. L'auteur propose que les apprenants, après avoir complété une tâche où ils se focalisent sur la forme, complètent une nouvelle tâche où le sens doit être utilisé. Newton (1995) a étudié huit apprenants réalisant quatre tâches communicatives. Aucune définition n'a été donnée par l'enseignant et aucun dictionnaire n'a été utilisé, ce qui fait que la source de l'acquisition était le contexte de l'activité et les savoirs des autres apprenants. En se focalisant sur l'apprenant qui a obtenu le score le plus bas lors du pré-test, Newton a trouvé que la négociation ne garantit pas l'acquisition car cinq des sept mots négociés avec succès n'ont pas été acquis. Il postule donc que c'est plutôt la tâche qui est le facteur le plus important pour l'acquisition. Si le mot est nécessaire pour la tâche, et si on doit comprendre le sens du mot, il y a de fortes chances qu'il soit appris. Newton a trouvé d'ailleurs une corrélation positive entre le nombre de rencontres d'un mot et acquisition.

La négociation facilite la compréhension, et la compréhension est nécessaire pour l'apprentissage du vocabulaire. Cependant, la compréhension ne garantit l'acquisition ni en présentiel (Ellis et al., 1994), ni en ligne (Smith, 2005). La tâche semble jouer un rôle principal. Dans notre corpus, chaque tâche est accompagnée d'un ou plusieurs mots-clés dont la compréhension est nécessaire pour compléter la tâche. Ce sont souvent ces mots-clés qui font l'objet de séquences d'explication lexicale.

2.4.5 La tâche

S'il est accepté que la négociation du sens facilite l'acquisition du vocabulaire et l'acquisition des langues en général, les questions qui se posent naturellement sont les suivantes : quels types de tâches suscitent le plus de négociation ? Quels types de tâches suscitent la négociation de meilleure qualité ? Certaines études ont trouvé que les tâches d'échanges d'informations suscitent le plus de négociation (Blake, 2000 ; Doughty & Pica, 1986 ; Ellis, 1991). Pour réaliser ce type de tâche qui se déroule souvent en binômes, chaque apprenant dispose d'informations que l'autre n'a pas. Les apprenants doivent communiquer entre eux afin d'obtenir les informations requises pour compléter la tâche. D'autres auteurs, comme Smith (2003), ont trouvé que les tâches de prise de décision suscitent plus de négociation. Dans ce type de tâche, les apprenants partagent leurs opinions personnelles afin d'arriver à un accord sur la meilleure façon de résoudre un problème.

D'autres études se focalisent sur la qualité de négociation plutôt que la quantité. Ainsi, Hill & Laufer (2003) ont comparé trois types de tâches et leur effet sur l'apprentissage : une tâche focalisée sur le message, une tâche réceptive focalisée sur la forme, et une tâche

productive sur la forme. Elles ont trouvé que les tâches qui se focalisent sur la forme sont plus efficaces pour la mémorisation à long-terme de mots que les tâches qui se focalisent sur le message. Pour ces auteures, ce n'est pas le temps passé sur la tâche qui compte, mais le temps passé sur l'item cible. Une tâche efficace est une tâche qui suscite le traitement profond des items cibles ainsi qu'une focalisation sur la forme.

Newton (2013) critique aussi des études qui ne focalisent que sur la quantité de négociation et ignorent la qualité. Pour pallier cette carence, il a analysé la façon dont des apprenants adultes d'anglais ont géré les mots inconnus qu'ils ont rencontrés lors de deux types de tâches communicatives : une tâche d'échange d'informations et une tâche d'échange d'opinion. Il a trouvé que quantitativement, les tâches d'échange d'informations ont suscité plus de deux fois plus de négociation, mais que qualitativement, la forme des mots inconnus a été négociée 20 fois plus que le sens pour ce type de tâche. Deux autres résultats de cette étude nous intéressent. Premièrement, Newton a trouvé que plus que la moitié des mots non-négociés ont été appris. Cela veut dire que même les items lexicaux qui ne font l'objet d'aucune séquence d'explication lexicale peuvent être appris par les apprenants. Newton postule que les mots qui sont contextualisés peuvent activer des savoirs qui ne seraient pas activés si le mot était décontextualisé. Deuxièmement, il a trouvé que tous les mots négociés ont été écrits avant d'être négociés, ce qui veut dire qu'il est plus facile de négliger un mot inconnu qui vient de la modalité orale. Nous répétons donc qu'il est important d'écrire les mots potentiellement inconnus, faute de quoi le risque de les ignorer est élevé.

Bogaards (1994 : 92) favorise aussi une évaluation qualitative des tâches et insiste que la difficulté de la tâche est l'un des facteurs les plus importants pour l'acquisition du vocabulaire. Il explique que « le degré de difficulté des décisions à prendre mène à des niveaux qualitativement différents de l'insertion des traces. En clair : les tâches difficiles mènent à des traces mémorielles mieux établies que les tâches faciles ». Un autre facteur déterminant est le degré d'implication personnelle de l'apprenant. Si l'apprenant est impliqué personnellement dans des tâches significatives, l'apprentissage sera favorisé. Cette implication personnelle est facilitée si la tâche a du contenu significatif et intéressant pour l'apprenant.

Il semblerait que la meilleure solution soit d'intégrer le travail sur le lexique dans des tâches. Les enseignants peinent souvent à enseigner du lexique en cours de FLE car ils sont préoccupés par la réalisation de tâches. Une bonne intégration de l'enseignement du lexique

s'impose, faute de quoi on aurait « l'impression que l'enseignement du vocabulaire est « plaqué » sur l'enseignement général de la langue » (Bourdon-Larger, 2008 : 83).

Dans notre corpus, plusieurs tâches sont proposées, et chacune s'accompagne d'un ou plusieurs mots-clés. Certaines tâches ont des objectifs concrets. Par exemple, lors de la quatrième semaine des interactions, les apprentis-enseignants doivent planifier un goûter d'anniversaire pour enfants dans un restaurant.[23] D'autres tâches nécessitent le partage d'opinions. Par exemple, lors de la première semaine, les apprentis-enseignants doivent donner leurs opinions sur les attitudes des travailleurs irlandais et sur la pause café. Ces tâches donnent naissance à de nombreuses séquences d'explication lexicale sur les mots-clés mais aussi sur des mots associés et sur des mots demandés par les apprenants.

2.4.6 Autonomie

Avant de conclure ce chapitre, il convient de rappeler que l'enseignant ne joue qu'un rôle très limité dans l'acquisition du vocabulaire. C'est plutôt la responsabilité de l'apprenant, ainsi que l'explique (Tréville, 2000 : 129) :

> « En matière de vocabulaire, plus qu'ailleurs peut-être, la tâche d'apprentissage semble être en grande partie du ressort de l'apprenant, de ses facultés innées (la mémoire, en particulier) et de sa détermination à poursuivre l'enrichissement jamais tout à fait achevé de son bagage lexical, par lui-même, au delà de la tutelle d'un programme éducationnel ».

L'autonomie de l'apprenant est d'autant plus indispensable qu'il est peu probable qu'un item lexical qui fait l'objet d'une explication soit connu et mémorisé suite à la séance, même s'il est compris. Bien qu'il soit impossible de déterminer un nombre exact ou un seuil (Huckin & Coady, 1999), Nation (2013) estime qu'il faut rencontrer un mot à peu près vingt fois dans des contextes divers avant de le connaître. Il faut aussi réviser les mots régulièrement en augmentant progressivement le temps entre chaque révision. C'est pourquoi il incombe aux apprenants de construire des listes de vocabulaire à réviser en dehors des cours. Plusieurs chercheurs (Folse, 2004 ; Galisson, 1983 ; Tréville, 2000) préconisent l'élaboration et l'usage d'un dictionnaire personnel. Dans le cahier envisagé par Folse (2004), qui aura assez d'espace blanc pour des révisions ultérieures, les apprenants mettront :

- l'item lexical ;
- une traduction ;
- un synonyme, antonyme, ou mot-clé qui relie sa forme et son sens ;
- une collocation courte.

[23] Une comparaison des explications de l'item « goûter d'anniversaire » sera proposée dans le chapitre 9.

Si on met chacun de ces éléments dans un coin, on peut favoriser le rappel en cachant un côté à la fois. Un autre avantage d'un dictionnaire élaboré par un apprenant est qu'il peut être organisé par thème au lieu de par ordre alphabétique, favorisant ainsi la création de réseaux sémantiques (Binon & Verlinde, 2004).

Plusieurs techniques de mémorisation existent, et la technique la plus préconisée est celle du mot-clé (Coady, 1993 ; Folse, 2004 ; Nation, 2013 ; Pressley, Levin, & McDaniel, 1987). Cette technique, qui nécessite une formation en amont et de l'entraînement, est utile surtout pour les mots dont la mémorisation est difficile. La technique consiste à penser à un mot en L1 qui ressemble phonologiquement au mot en L2.[24] Puis, on construit une image mentale qui relie les deux choses. Pour donner un exemple qui vient de nos souvenirs du collège, le mot *manger* en français se ressemble phonologiquement au mot *manger* en anglais (mangeoire). L'image mentale que nous avons formée consistait en une vache en train de manger dans une mangeoire. Il a été démontré expérimentalement que cette technique mnémonique est plus efficace que l'apprentissage par cœur (la répétition), l'usage d'images, l'usage du contexte, et l'usage de synonymes. Nation (2013 : 467) résume les principales raisons scientifiques pour lesquelles cette technique est si efficace pour la mémorisation d'items lexicaux :

1. Selon le principe des niveaux de traitement, plus une chose est traitée profondément et pensivement, mieux elle est retenue ;
2. Selon le principe de la focalisation d'attention, on apprend ce sur quoi on se focalise ;
3. Selon le principe d'attention délibérée, l'apprentissage délibéré amène à un apprentissage rapide et durable ;
4. Selon le principe de double encodage, la présence d'associations linguistiques et visuelles aide à la mémorisation.

Bien que de telles techniques existent, nous sommes d'accord avec Folse (2004) que les stratégies d'apprentissage ne constituent pas une raison pour faire l'économie de l'enseignement explicite du vocabulaire.

2.5 Conclusion

L'enseignement systématique et explicite du vocabulaire est nécessaire en langue étrangère, principalement pour les niveaux débutant et intermédiaire. Bien que l'apprentissage

[24] Quand il n'y a aucun mot qui ressemble au mot qu'on souhaite apprendre, la répétition reste la meilleure solution pour forger une représentation phonologique mentale.

incident soit le mécanisme principal par lequel les connaissances lexicales sont acquises en langue première, celui-ci est trop lent et problématique pour être la source principale du nouveau vocabulaire en langue étrangère. Il revient donc aux enseignants d'enseigner rapidement, de manière explicite et systématique, les mots de haute fréquence et les mots dont les apprenants auront besoin pour communiquer dans leurs domaines professionnels respectifs.

Un item lexical peut être constitué d'un seul mot graphique ou d'une suite de mots graphiques (unité polylexicale). Nous utilisons les termes *mot* et *item lexical* pour ne pas emprunter la notion d'*unité lexicale* proposée par Bogaards (1994) qui ne prend en compte qu'une seule acception d'un mot à la fois. La connaissance d'un item lexical est pluridimensionnelle : au plus simple, elle englobe la forme, le sens, et le usages.

Il n'est pas réaliste de s'attendre à ce qu'un item lexical soit entièrement connu suite à une seule explication. Il n'est pas réaliste non plus de supposer que si un mot est compris suite à une explication, il est appris. Cependant, il y a plusieurs facteurs qui facilitent l'apprentissage d'un item lexical. Ceux-ci méritent d'être privilégiés :

Pratiques facilitatrices	**Références**
Le repérage des formes orales et écrites ainsi qu'une focalisation sur celles-ci	(García, 2008 ; Groot, 2000 ; Laufer, 1994 ; Marijanović et al., 2008 ; Nation & Webb, 2011 ; Nation, 2013 ; Schmidt, 1990 ; Schmitt, 2008 ; Tréville, 2000)
Une focalisation sur le lien entre la forme et les sens	(Baddeley, 1993 ; Bogaards, 1994 ; Nation, 2013)
La contextualisation des items lexicaux	(Laufer & Shmueli, 1997 ; Tréville, 2000 ; Tréville & Duquette, 1996)
L'apprentissage de mots en groupes thématiques et non en groupes sémantiques	(Erten & Tekin, 2008 ; Nation, 2000 ; Théophanous, 2008; Tinkham, 1997)
L'usage de gloses et d'explications simples, y compris des traductions	(Brett et al., 1996 ; Chaudron, 1982 ; Ellis, 1995 ; Folse, 2004 ; Hulstijn et al., 1996 ; Kelly & Bruen, 2015 ; Laufer & Shmueli, 1997 ; Nation, 2013 ; Sweller, 2005b)
L'usage de dictionnaires Plusieurs rappels réceptifs et productifs	(Baddeley, 1993 ; Folse, 2004 ; Hulstijn et al., 1996 ; Nation & Webb, 2011)
L'usage génératif	(Fasel Lauzon, 2014 ; Joe, 1995, 1998)
La négociation du sens	(Ellis et al., 1994 ; Hatch, 1978 ; Long, 1981, 1983 ; Newton, 1995 ; Pica, 1994 ; Pica, Young, & Doughty, 1987 ; Porquier & Py, 2008 ; Swain, 1985, 1995)

La participation active ou du moins l'anticipation de participer activement	(Brouwer, 2003 ; Dobinson, 2001 ; Ellis et al., 1994 ; Fasel Lauzon, 2014 ; Laufer & Hulstijn, 2001 ; J Newton, 2001 ; Newton, 1995 ; Stahl & Clark, 1987 ; Stahl & Vancil, 1986)
La mise en place de tâches difficiles qui provoquent la négociation du sens et dont la réalisation nécessite une compréhension suffisante des mots-clés	(Blake, 2000 ; Bogaards, 1994 ; Bourdon-Larger, 2008 ; Doughty & Pica, 1986 ; R. Ellis, 1991 ; Hill & Laufer, 2003 ; Newton, 2013 ; Smith, 2003)
L'élaboration de dictionnaires personnels et la révision de ceux-ci en dehors du cours	(Binon & Verlinde, 2004 ; Folse, 2004; Galisson, 1983 ; Tréville, 2000)
L'usage de techniques mnémoniques comme celle du mot-clé	(Folse, 2004 ; Nation, 2013 ; Pressley et al., 1987)

Figure 4 : Pratiques favorisant l'apprentissage du vocabulaire

La façon dont les apprentis-enseignants mettent en place les conditions favorables à l'acquisition du lexique en visioconférence reste à analyser, et les facteurs listés ci-dessus seront pris en compte lors de nos analyses des séquences d'explication. Compte tenu de l'importance de ces facteurs, nous nous intéressons lors de nos analyses à :

1. La façon dont les apprentis-enseignants se focalisent sur la forme des items lexicaux (chapitre 6) ;
2. La façon dont les apprentis-enseignants se focalisent sur le sens des items lexicaux (chapitre 7) ;
3. La façon dont les apprentis-enseignants impliquent les apprenants dans les séquences d'explication lexicale (chapitre 8).

Un autre facteur que favorise l'acquisition du vocabulaire et qui joue un rôle important dans des séquences de négociation du sens, surtout lors des séquences d'explication lexicale, est l'encodage multimodal. C'est ce point qui sera traité dans le chapitre suivant.

Chapitre 3 : Vers une approche multimodale de l'analyse des séquences d'explication lexicale en visioconférence

3.1 Introduction

Nous avons évoqué à la fin du chapitre précédent que l'encodage multimodal favorise la création de traces mémorielles. En effet, la multimodalité joue un rôle important non seulement pour l'acquisition du vocabulaire mais aussi pour la communication, que ce soit en visioconférence ou en présentiel. L'objectif de ce chapitre est d'argumenter en faveur d'une approche multimodale de l'analyse des interactions pédagogiques en visioconférence, et plus particulièrement des séquences d'explication lexicale. Nous ferons donc une courte revue de la littérature existante sur les séquences de négociation du sens et des explications lexicales en faisant ressortir l'absence quasi-totale de prise en compte de la multimodalité dans les études sur la négociation du sens et sur les explications lexicales. Puis, nous nous appuierons sur l'importance de la multimodalité de la communication afin de proposer une conception multimodale des séquences d'explication lexicale.

Comme le rappellent (O'Halloran & Smith, 2011), il n'existe pas une théorie unique de la multimodalité. Ces auteurs expliquent qu'il y a deux types de champs disciplinaires de la multimodalité. Il y a d'une part des études qui appliquent des théories et méthodes existantes pour étudier des phénomènes multimodaux et d'autre part il y a des études qui traitent des enjeux uniques de la multimodalité comme la modélisation de ressources sémiotiques autres que le langage verbal oral, et le traitement d'enjeux méthodologiques comme la transcription, l'analyse et la représentation. Il y a aussi des approches qui combinent ces deux types d'études. Dans ce chapitre nous allons nous appuyer sur des théories déjà développées par des chercheurs étudiant la négociation du sens et le langage corporel (études gestuelles) afin d'aborder certaines notions proprement multimodales comme la définition de modes sémiotiques.

3.2 Une conception classique des séquences de négociation du sens

Schegloff, Jefferson et Sacks (1977) ont été parmi les premiers à s'intéresser aux problèmes de communication liés au code. Dans un article séminal, ces auteurs se sont intéressés aux séquences de réparation qui surgissent dans la conversation ordinaire. Ces séquences de réparation sont nécessaires pour que la conversation puisse continuer à se dérouler au cas où il y a une incompréhension, ce qui arrive assez souvent. Pour ces auteurs, une réparation consiste à remplacer une erreur (dite *trouble source*) par ce qui est correct. Ils notent qu'une réparation peut avoir lieu même si l'erreur n'est pas audible. A l'aide d'un exemple tiré de notre corpus, nous allons analyser une réparation.

Transcription de la parole		
1	ALN	en irlande/ euh: (1,0) on doit euh retirer à: euh: à soixante (.) soixante-cinq ans/ (2,3)
2	ADL	soixante-cinq ans\ (0,9) alors retirer/ (0,9) retirer/ (0,5) c'est retraite\ (0,3) retraite\ (0,2) regarde\ <((clavardage)) retraite> la retraite\

Exemple 1 : Une hétéro-réparation auto-initiée

A la ligne 1, Alannah dit « retirer » au lieu de « retraite ». Le mot « retirer » est donc le *trouble source*. A la ligne 2, Adèle corriger son interlocutrice. On assiste donc à une *hétéro-correction*. Cette hétéro-correction est *auto-initiée* car Adèle prend l'initiative de corriger son interlocutrice sans qu'on le lui demande.

Schegloff, Jefferson et Sacks ont distingué entre auto- et hétéro-réparations, celles-ci étant auto- ou hétéro-initiées. Une auto-réparation a lieu lorsque l'énonciateur du *trouble source* s'auto corrige. Il y a hétéro-réparation, à l'inverse, quand le récepteur du *trouble source* corrige l'énonciateur du *trouble source* (cf. Exemple 1). Une réparation est auto-initiée lorsque celui qui corrige le *trouble source* initie la réparation lui-même (cf. Exemple 1), que ce soit une auto- ou une hétéro-réparation. Inversement, une réparation est hétéro-initiée lorsque celui qui corrige le *trouble source* est invité à le faire par son interlocuteur. De même, une réparation hétéro-initiée peut aboutir à une hétéro- ou auto-réparation. Le tableau ci-dessous résume les quatre types de réparations que nous venons de décrire.

	Auto-initiée	Hétéro-initiée
Auto-réparation	L'énonciateur du trouble source corrige son propre énoncé sans qu'on le lui demande.	L'énonciateur du trouble source corrige son propre énoncé suite à une demande de son interlocuteur.
Hétéro-réparation	Le récepteur du *trouble source* prend l'initiative de corriger son interlocuteur (cf. Exemple 1).	Le récepteur du *trouble source* corrige l'énoncé de son interlocuteur à la suite d'une demande de correction.

Figure 1 : Les réparations (Schegloff, Jefferson, & Sacks, 1977)

Les auteurs notent que les hétéro-initiations aboutissent massivement à des auto-réparations. Finalement, pour ces auteurs, ce sont les auto-réparations qui sont largement préférées par les interlocuteurs, les hétéro-réparations étant beaucoup plus rares dans la conversation ordinaire. Les auteurs notent toutefois une exception remarquable : quand il s'agit d'interactions entre adulte-enfant, ou d'interactions natif-non-natif, les hétéro-réparations sont tolérées. Les auteurs, bien que parlant de « adulte-enfant » et non de « enseignant-apprenant », ont pris le soin de préciser que lorsqu'il y a un déséquilibre entre les compétences langagières des interlocuteurs, la fréquence d'hétéro-réparations augmente. Selon eux, cette situation est temporaire et il s'agit d'une étape intermédiaire avant que le locuteur moins compétent ne parvienne à s'auto-corriger. Les travaux sur la réparation entrepris par ces pionniers du domaine de l'analyse de la conversation ont donné naissance à maintes études sur la négociation du sens.

La communication exolingue a été définie par (Porquier, 1979 : 50) comme « celle qui s'établit entre individus ne disposant pas d'une L1 commune, et qui représente, dans des situations extrêmement diverses, une partie importante des échanges humains » (1979 : 50). Ultérieurement, la définition a été approfondie par cet auteur (1984 : 18) comme l'interaction dont les caractéristiques principales sont les suivantes :

> « les participants ne peuvent ou ne veulent communiquer dans une langue maternelle commune (...)
>
> - les participants sont conscients de cet état de choses ;
>
> - la communication exolingue est structurée pragmatiquement et formellement par cet état de choses et donc par la conscience et les représentations qu'en ont les participants ;
>
> - les participants sont, à divers degrés, conscients de cette spécificité de la situation et y adaptent leur comportement et conduites langagières ».

On parle donc de conversation exolingue quand les asymétries de code ne peuvent pas être ignorées par les interlocuteurs. Lors de la conversation exolingue, plus un interlocuteur non natif est impliqué dans l'interaction, et moins il partage une langue et un niveau de compétence, plus il y aura de séquences de négociation (Varonis et Gass, 1985). Les

séquences de négociation du sens surviennent principalement lorsqu'on est en train de réaliser des tâches (Nicolaev, 2010) et lorsqu'un contrat didactique[1] se met en place. Nous utilisons la définition de contrat didactique proposée par De Piétro, Matthey et Py (1989 : 8) où

> « le contrat didactique dans la conversation exolingue ne peut être instauré que par une définition commune de la situation. L'asymétrie de la situation est acceptée et la complémentarité des rôles reconnue: le natif peut et doit enseigner sa langue, l'alloglotte doit donner des quittances du savoir qu'il reçoit, sous la forme généralement d'une répétition de la proposition du natif ».

Ces auteurs rappellent d'ailleurs que quand le contrat didactique est non-demandé, non-explicite, non-négocié ou n'est pas compris de la même façon par l'alloglotte et le natif, des difficultés relationnelles, et un certain malaise peuvent survenir.

Ces séquences de négociation nous intéressent dans le domaine de l'acquisition des langues étrangères car selon Long (1991), les séquences propices à l'acquisition se focalisent sur la forme. Nicolaev (2010 : 171) résume ainsi la focalisation sur la forme :

> « La focalisation sur la forme implique donc des processus cognitifs permettant à l'apprenant de porter son attention sur le code linguistique en même temps qu'il s'implique dans une situation de communication. Ces moments de focalisation se concrétisent à travers les négociations de sens. Ces dernières émergent lorsque la structure discursive normale de l'interaction est modifiée suite à une demande de répétition, une demande de clarification, une vérification de compréhension ou de reformulation du matériau langagier (input) ».

Afin de mieux appréhender tout ce qui se passe lors des séquences de négociation, il convient d'en avoir un modèle.

3.2.1 Modéliser les séquences de négociation du sens

Les séquences de négociation du sens peuvent arriver à tout moment de l'interaction et pour des raisons diverses. Elles prennent la forme de séquences latérales (Jefferson, 1972) où la conversation s'arrête pour passer à un autre sujet de discussion. Lors de l'enseignement des langues par visioconférence, les séquences latérales émergent souvent pour traiter des problèmes liés au langage (Rivens Mompean & Cappellini, 2015). Un des modèles les plus cités est celui de Varonis et Gass. Ces auteurs (1985 : 3) définissent les séquences d'incompréhension comme « les échanges dans lesquels il y a une indication manifeste que la compréhension entre participants n'a pas été complète ».[2] Leur modèle est constitué de deux grandes parties : le déclencheur (« *trigger* ») et la résolution.

[1] « Le concept de contrat didactique a été introduit en didactique des mathématiques par G. Brousseau qui l'a défini comme « ces habitudes (spécifiques) du maître attendues par l'élève et ces comportements de l'élève attendus par le maître » (1980, p. 181) » (cité par De Pietro, Matthey et Py, 1989 : 6).

[2] « those exchanges in which there is some overt indication that understanding between participants has not been complete » (Varonis & Gass, 1985 : 3).

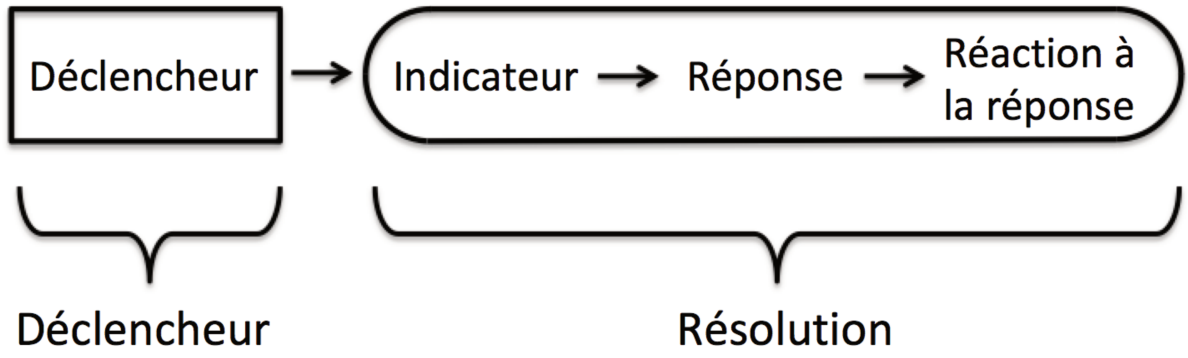


Figure 2 : Modèle de négociation (Varonis & Gass, 1985 : 74)

Le déclencheur est l'énoncé ou partie de l'énoncé qui n'est pas compris par l'interlocuteur, ce dernier pouvant l'ignorer ou réagir. S'il l'ignore, soit il espère que l'énonciateur lui donnera plus d'informations, soit il considère qu'il est sans importance. Si un déclencheur reste ignoré, il est souvent impossible de savoir s'il y a eu incompréhension. Dans ce cas, celui-ci demeure un déclencheur potentiel.

L'indicateur est l'énoncé ou partie de l'énoncé qui indique qu'il y a incompréhension. L'indicateur arrête le déroulement de l'interaction pour traiter l'élément problématique. Les indicateurs peuvent être implicites ou explicites et peuvent se manifester de plusieurs manières différentes. Schegloff, Jefferson et Sacks (1977) parlent de *techniques d'initiation*, qui diffèrent selon que la séquence est auto ou hétéro-initiée. Les réparations auto-initiées arrivent dans le même tour de parole que le déclencheur (*trouble source* selon les auteurs) et ont comme initiateurs des perturbations non-lexicales comme des arrêts abrupts, amorces de mots, prolongations de mots, marqueurs d'hésitation, etc. Les réparations hétéro-initiées utilisent un groupe d'unités de construction du tour (UCT) pour initier la réparation, comme « hein ? » « quoi ? » « qui ? » etc. D'autres stratégies incluent la répétition d'une partie du tour précédent, suivi ou non d'un marqueur de questions, ou bien la reformulation du tour précédent précédé par « vous voulez dire ». Un indicateur peut aussi être une vérification de compréhension. Cette liste de stratégies n'est pas exhaustive. Si l'émetteur du déclencheur ne répond pas à l'indicateur, soit il ignore qu'une réparation est requise, soit il refuse de coopérer. Notons que toutes ces stratégies sont verbales et ne prennent pas en compte la dimension multimodale de la parole.

La « réponse » est la réponse à l'indicateur qui prend en compte l'incompréhension. Schegloff, Jefferson et Sacks notent que si l'indicateur est vague du type « je ne comprends pas », le déclencheur n'est pas localisé et la réponse consiste à tout répéter. Smith (2003) liste plusieurs types de réponses : réponse minimale, répétition du déclencheur avec ou sans

modification de la syntaxe, déclaration d'une incapacité à répondre, reformulation ou élaboration de l'élément problématique. Ces réponses sont catégorisées comme étant des réponses minimales, modifiées ou élaborées. Une réponse peut témoigner d'un certain niveau d'uptake[3]. Or, la présence ou absence d'uptake pendant les activités en ligne ne semble pas être corrélée à l'acquisition du lexique (Smith, 2005).

La quatrième étape, la réaction à la réponse (RR), est facultative. Durant cette étape, le destinateur de la réponse peut réagir. La RR peut être minimale du type « ok » ou « oui je comprends ». Smith (2003) ajoute que l'on peut donner une RR du type métalinguistique où on commente la provenance du déclencheur.

Le modèle proposé par Varonis et Gass est très généraliste et n'a pas été conçu spécifiquement pour l'interaction en ligne. Smith (2003) a repris le modèle de Varonis et Gass et l'a élaboré spécifiquement pour l'interaction synchrone à distance (clavardage) en se basant sur un corpus d'interaction. Il a ajouté deux nouvelles étapes qui prolongent le modèle de Varonis et Gass. L'étape de « confirmation » (C) et de « reconfirmation » (RC). L'étape de confirmation peut être positive ou négative. Si elle est négative, celui qui a donné la réponse (ou l'explication) donne plus d'input[4] ou, rarement, abandonne la séquence. Si elle est positive, celui qui a fourni la réponse a recours à trois types de confirmation : confirmation simple, réaffirmation, et validation de compréhension. La reconfirmation est la dernière étape, facultative, qui consiste normalement en une réponse minimale du type « ok », « bon », « oui ». La présence quelque peu élevée de cette étape lors de la communication médiée par ordinateur est due au besoin de réaffirmation explicite en ligne. Il convient de répéter encore une fois que la présence d'une réaction à la réponse, ou d'une reconfirmation, ne garantit pas que l'apprenant a appris l'élément langagier qui fait l'objet de la séquence de négociation (Smith, 2005).

Un nouveau modèle avec de nouvelles étapes a été nécessaire pour la communication par clavardage car ce type de communication est différent de la communication en présentiel[5]. En effet, la modalité écrite exclut certains indices prosodiques et kinésiques dont la conversation orale est largement pourvue. Par exemple, on ne peut pas affirmer sa compréhension en faisant des hochements de tête par clavardage : il faut écrire « oui, j'ai

[3] Autrement appelé « saisie », le terme « uptake » fait référence à ce qui est compris, et éventuellement appris, par l'apprenant (Cuq, 2003).

[4] Autrement appelé « exposition », le terme « input » fait référence à l'environnement langagier de l'apprenant, et ce à quoi il est exposé (Cuq, 2003). Input ne garantit pas uptake (Smith, 2005).

[5] Nicolaev (2010) a trouvé que le modèle élaboré de Smith est pertinent pour la visioconférence car dans son étude, 62% des séquences de négociation du sens observées étaient étendues dans le sens où elles comportaient les étapes supplémentaires de Smith.

compris ». Ainsi, les étapes de confirmation et de réaffirmation proposées par Smith deviennent utiles pour la négociation du sens. Des indices visuels qui sont absents lors de la communication par clavardage redeviennent en partie visibles en visioconférence, ce qui rend une approche multimodale encore plus importante. La visioconférence étant un contexte qui semble être propice à la négociation du sens (Nicolaev, 2010; Sarré, 2011), l'étude de la multimodalité lors des séquence de négociation du sens en visioconférence est digne d'intérêt.

3.2.2 Modéliser les séquences d'explication lexicale

Nous venons de décrire un modèle classique pour tout type de séquence de négociation du sens, y compris les séquences qui émergent de la conversation ordinaire. Nous allons nous focaliser dans cette section sur les interactions didactiques, particulièrement sur les séquences d'explication lexicale. Une séquence d'explication lexicale est une séquence de négociation du sens qui s'impose lorsque c'est un item lexical qui fait obstacle à la communication. Autrement dit, le *trouble source* est un item lexical. Rappelons que la plupart des déclencheurs sont de type lexical aussi bien en présentiel (de Pietro, Matthey, & Py, 1989; Fasel Lauzon, 2014 ; Williams, 1999) qu'en ligne (Blake, 2000 ; McDonough & Sunitham, 2009 ; Nicolaev, 2010 ; Smith, 2003). Quand un item lexical provoque une incompréhension c'est un « moment de crise » qui est « apte à engendrer – et à rendre manifeste – un travail lexical » (Lüdi, 1991 : 200). Un obstacle lexical surgit lorsque « l'information lexicale mémorisée par les interactants n'est pas identique », et devient alors une « tâche communicative commune » (Lüdi, 1991 : 201). Lors de cette tâche commune, il y a un « travail lexical » et une « synchronisation codique » (Lüdi, 1991 : 197) qui consiste en « un sous-ensemble d'opérations discursives de négociation visant à obtenir un accord sur le code employé ». C'est sur ce travail langagier visant à résoudre une incompréhension liée au lexique que nous allons nous focaliser lors de nos analyses.

L'exemple suivant montre que le modèle proposé par Varonis et Gass peut être utilisé pour analyser une séquence d'explication lexicale.

Parole orale			Image de Victor	Phénomènes observés
1	VIC	est-ce que tu comprends l'expression (.) on veut les trente-cinq heures (1,9)		Geste métaphorique représentant le concept des trente-cinq heures
2	LIM	hum: (3,0) euh: (0,4) il veut euh beaucoup des heures/		
3	VIC	(1,0) d'accord/ (0,5) en fait/ (.) en france/ (0,2) euh selon la loi/ les gens ne peuvent pas travailler/ (0,2) plus que trente-cinq heures\ (0,5) donc **trente-cinq heures c'est le maximum** prévu par la loi\ (0,6)		
4	LIM	oui je comprends		

Exemple 2 : Une séquence d'explication lexicale

A la ligne 1, le *déclencheur* est l'expression « on veut les trente-cinq heures », car la réponse incorrecte de l'apprenant à la ligne 2, l'*indicateur*, montre que l'expression n'a pas été comprise. Victor donne la *réponse* à la ligne 3 quand il explique le sens des trente-cinq heures à l'aide d'un geste métaphorique, et Liam donne une *réaction à la réponse* à la ligne 4.

Les séquences d'explication lexicale surviennent souvent lors de l'interaction pédagogique et peuvent êtres de deux types différents :

1. Un interlocuteur (souvent l'apprenant) éprouve des difficultés à comprendre un item lexical qu'il rencontre. Il s'agit donc d'un problème de décodage. Dans ce cas, l'enseignant ou un autre apprenant doit *faire comprendre* une ou plusieurs facettes de la connaissance de l'item lexical[6] à celui qui ne comprend pas. L'item lexical fait donc l'objet d'une explication (cf. Exemple 2 ci-dessus).
2. Au cours de la production langagière, un item lexical fait défaut à un apprenant et l'empêche de s'exprimer convenablement. Il s'agit donc d'un problème d'encodage. Un travail collaboratif s'impose afin de trouver ce que Grossmann (2011) appelle le *mot juste*. L'apprenant a souvent besoin d'expliquer le sens de l'item lexical recherché afin que ses interlocuteurs puissent le lui donner ou lui faire des propositions.

Dans les deux cas, une explication s'impose. On s'intéresse aux explications d'autant plus que « la compétence pédagogique s'identifie à la compétence explicative. Et, symétriquement, le

[6] La forme, le sens, et l'usage (cf. chapitre précédent).

professeur ne peut évaluer l'effet de son enseignement sur l'apprentissage des étudiants qu'en les mettant en situation d'avoir à expliquer » (De Gaulmyn, 1991 : 279).

Malgré cette distinction, Lüdi (1991) explique que le travail sur l'encodage et le travail sur le décodage sont parfois reliés et qu'il n'y a pas une forte distinction entre les deux. Si, par exemple, faute de connaissances lexicales chez l'apprenant, l'enseignant tente de parler à la place de celui-ci qui éprouve des difficultés de production langagière (hétéro-achèvement) et que ceci engendre une difficulté de compréhension, la séquence sera alors liée à la fois à une difficulté de production et à une difficulté de compréhension. Cette distinction est d'autant plus opaque que pour le lexique, les compétences réceptives et productives se chevauchent et sont interdépendantes (Nation & Webb, 2011 ; Nation, 2013).

Miecznikowski (2005 : 21) donne une définition générale d'une séquence d'explication lexicale qu'elle qualifie de *séquence métalexicale* :

> « a) Les séquences métalexicales sont des séquences interactionnelles pendant lesquelles des locuteurs se focalisent réflexivement sur une unité lexicale du discours en train de se faire.
>
> b) Les séquences métalexicales sont des séquences qui sont consacrées au traitement d'un problème lié à l'unité lexicale en question. Ce *problème lexical* est soulevé en début de séquence et traité interactivement jusqu'à ce que les participants arrivent à un accord local qui leur permet de clore la séquence ».

Cette définition très générale peut être complétée par d'autres chercheurs qui décrivent plus finement le déroulement d'une séquence d'explication lexicale en la divisant en phases. La structure canonique à laquelle nous souscrivons a été modélisée par Gülich (1990a), De Gaulmyn (1991) et Lüdi (1991) et reprise par Fasel Lauzon (2008, 2014). Ce modèle est constitué de trois étapes :

I. L'ouverture : l'item lexical non compris est identifié, repéré, localisé, mis en exergue, et problématisé.

II. Le noyau explicatif : l'item lexical non compris est expliqué, le plus souvent par l'enseignant, mais souvent en collaboration avec les apprenants. Parfois, le noyau est pris en charge par les apprenants. Dans le cas d'une recherche lexicale, le mot juste est donné et/ou expliqué.

III. La clôture : l'enseignant ratifie et évalue la compréhension de l'apprenant. « Les partenaires s'accordent pour confirmer la réparation apportée au problème, ce qui clôt la séquence explicative » (De Gaulmyn, 1991 : 284).

Ces trois phases s'alignent avec les « trois constituants macro-structurels du type textuel explicatif : questionnement - problématisation ; résolution ; conclusion » de Coltier (1986) (cité par Ducancel, 1991 : 120). Une séquence d'explication lexicale peut être constitué de ces

trois phases, mais peut aussi n'en contenir qu'une ou deux. Nous avons suivi ce modèle canonique pour diviser les séquences d'explication lexicale de notre corpus en phases. Ce repérage sera discuté dans le chapitre 5.

Notons que ce modèle peut être appliqué à l'exemple 2 ci-dessus : dans l'ouverture, l'item lexical est repéré et problématisé (lignes 1-2). Dans le noyau, il est expliqué (ligne 3), et dans la clôture (ligne 4), l'apprenant indique qu'il a compris.

3.2.3 Rôles des participants

Selon De Gaulmyn (1991 : 282-283), une explication réunit deux interlocuteurs autour d'un objet à comprendre. Elle donne des étiquettes aux participants humains et à l'objet :

> « – **S2**, un sujet qui ne comprend pas, qui veut comprendre, qui risque de ne pas comprendre, le sujet auquel s'adresse l'explication et/ou qui la reçoit, demandeur, destinataire, récepteur, soit l'**explicataire**.
>
> – **S1**, un sujet qui explique, un expert ou qui en fait fonction, l'**explicateur**.
>
> – **O**, un objet à comprendre, objet problématique, l'**explicable**, ancrage ou occasion de l'activité explicative ».

Il convient de préciser ici que les rôles de S1 ou S2 ne sont pas figés, et ils peuvent être joués par l'apprenant ou par l'enseignant. Dans le cas où l'apprenant ne comprend pas une unité lexicale O, celui-ci jouerait le rôle de S2 et l'enseignant jouerait le rôle de S1. Dans le cas d'une séquence liée à un problème de production, l'apprenant peut jouer le rôle de S1 qui explique le sens de O à l'enseignant S2 afin que ce dernier puisse fournir le nom de O en français. Il peut y avoir une alternance des rôles à l'intérieur d'une même séquence. Par exemple, dans cette même séquence, l'enseignant, après avoir donné l'item lexical recherché à l'apprenant, peut passer de S2 à S1 afin de fournir plus d'informations lors de la phase de confirmation (Smith, 2003). Nous verrons dans le chapitre 8 que les apprenants sont souvent invités par les apprentis-enseignants à jouer le rôle d'explicateur.

Nous avons résumé une façon très générale de conceptualiser et de modéliser les séquences de négociation du sens ainsi que les séquences d'explication lexicale. Nous avons appliqué ces modèles à un exemple. Ces modélisations ne sont pas intrinsèquement monomodales ou multimodales. Cependant, aucune de ces études ne prend en compte le caractère multimodal des séquences de négociation du sens ou d'explications lexicales. Par exemple, Miecznikowski (2005 : 46) décrit différentes façons d'initier une réparation, exclusivement verbales : *qui ?*: *que ?*: *quoi ?*: *lequel ?*, « une répétition partielle du tour avec remplacement de l'élément problématique par un pronom interrogatif », « une répétition partielle du tour », ou une expression du style « vous vouliez dire ». Tous ces initiateurs

linguistiques ne sont véhiculés que par un seul mode : le voco-verbal. Quand cette auteure dit « un geste », c'est l'une des rares fois où elle mentionne l'usage de ressources sémiotiques autre que le verbal, mais le geste n'est pas représenté visuellement et se voit accorder une place secondaire et périphérique dans la hiérarchie des modes. D'autres auteurs comme Lazaraton (2004) déplorent la carence de mentions du langage non-verbal dans la littérature sur la linguistique, celui-ci étant perçu comme une ressource compensatoire et accessoire. Or, lors d'une séquence de négociation du sens, les indices multimodaux comme la prosodie, les gestes, les expressions faciales, et changement de posture jouent un grand rôle non seulement dans l'indication d'une incompréhension (Develotte & Drissi, 2013 ; Nicolaev, 2010) mais également dans la séquence de négociation où l'incompréhension est résolue.

3.3 Vers une perspective multimodale

Nous adoptons une approche multimodale pour analyser les explications lexicales en visioconférence. Puisque le dispositif Visu est multicanal (son, vidéo et clavardage), le caractère multimodal des échanges que nous observons est évident, même si multicanalité et multimodalité ne sont pas synonymes. Il est possible d'argumenter que toutes les interactions, qu'elles soient médiées ou non, sont *a priori* multimodales (Norris, 2004). Une conversation téléphonique est multimodale dans le sens où le volume, le timbre, le débit, et l'intonation de la voix sont variables et créent du sens. Dans une perspective multimodale, chaque ressource socio-sémiotique a pour vocation de contribuer à la production du sens (Sindoni, 2013). Selon Bateman (2011), la tâche centrale de la recherche multimodale consiste à démontrer qu'une combinaison de modes accroît la possibilité d'expression par rapport à un seul mode. Nous allons argumenter en faveur d'une approche multimodale pour analyser les séquences d'explication lexicale en visioconférence. Nous allons définir la multimodalité ainsi que les modes sémiotiques qui nous aideront par la suite à analyser les séquences d'explication lexicale.

3.3.1 Définir la multimodalité

Dans le cadre d'une étude sur l'enseignement du FLE en visioconférence, Develotte, Guichon, & Kern (2008 : 132) ont défini la multimodalité « comme englobant, d'une part, les dimensions verbale, non verbale et paraverbale (voix, débit, prosodie, gestes, mimiques, rires) et, d'autre part, la co-présence de différents outils de communication avec la juxtaposition possible d'écriture, de parole, d'image ». Cette définition de la multimodalité reprend les

dimensions humaines et techniques, notions qui avaient déjà été mises au point avant l'émergence de l'enseignement du FLE en visioconférence. Il convient donc de revisiter ces notions de base et de retracer le chemin entre la multimodalité de la communication et son application à l'interaction en visioconférence. Kress & Van Leeuwen (2001) ont été parmi les premiers à dire qu'un texte n'est pas seulement linguistique. Un texte peut comprendre également des animations, du son, et des vidéos. Les enregistrements vidéo dont nous disposons constituent donc des textes multimodaux selon cette définition. La multimodalité prend en compte non seulement le langage mais aussi s'intéresse à toute la gamme de ressources sémiotiques qui sont utilisées pour communiquer. Jewitt (2014a) explique qu'une approche multimodale comprend quatre présuppositions théoriques :

1. Le langage fait partie d'un ensemble multimodal. Ce n'est pas uniquement le langage qui compte et qui est toujours au centre de l'évènement communicatif.
2. Chaque mode réalise une tâche communicative différente à chaque instant. Ces tâches ne sont pas figées, mais sont articulées et situées.
3. On communique en sélectionnant et en configurant les modes.
4. Le sens qui est créé par les ressources sémiotiques est social. Il suit les normes et les règles sociales.

La multimodalité a été définie par Kress & Van Leeuwen (2001 : 20) comme

> « l'usage de plusieurs modes sémiotiques dans la conception d'un produit ou évènement sémiotique, avec la façon particulière dont ces modes sont combinés — les modes peuvent par exemple se renforcer les uns les autres ('dire la même chose de façons différentes'), exercer des rôles complémentaires [...], ou être organisés hiérarchiquement [...] » (notre traduction)[7].

Cette définition aborde la notion de *mode* ainsi que l'organisation des modes. Nous allons donc définir les ressources sémiotiques et modes avant de parler de leur structuration.

3.3.1.1 Ressources sémiotiques

Les ressources sémiotiques ont été définies par Van Leeuwen (2005) et par (Jewitt, 2014a) comme les actions, les matériels et les artefacts que l'on utilise pour communiquer. Les ressources peuvent être nos cordes vocales, nos muscles, ou des technologies inventées par l'homme, comme le stylo ou l'ordinateur. Les ressources sémiotiques englobent d'une part les supports physiques que l'on utilise pour communiquer, et d'autre part les actions par lesquelles on communique : parler, écrire, gestualiser, etc.

[7] "the use of several semiotic modes in the design of a semiotic product or event, together with the particular way in which these modes are combined – they may for instance reinforce each other ('say the same thing in different ways'), fulfill complementary roles […], or be hierarchically ordered […]" (Kress & Van Leeuwen, 2001 : 20).

3.3.1.2 Mode

La notion de *mode* renvoie à la façon dont une ressource sémiotique est formée culturellement par son usage (Jewitt, 2014a ; Kress, 2014). Kress (2014) donne les exemples suivants de mode : images, écriture, disposition, musique, gestes, paroles. Ceux-ci sont utilisés en communication et leurs usages obéissent à des normes sociales. Pour Kress, l'aspect social est primordial : qu'est-ce qui est considéré comme un mode par un groupe social. Par exemple, la communauté francophone est capable de dire si quelqu'un est en train de *parler français* ou non.

3.3.1.3 Mode versus médium

Kress & Van Leeuwen (2001) opèrent une distinction entre *mode* et *médium*. Pour ces auteurs, un *mode* est constitué de ressources sémiotiques qui permettent de réaliser un discours et d'interagir. Un *médium* est un ensemble de ressources matérielles qui sont utilisées pour produire un évènement sémiotique. La plupart du temps, ces ressources sont fabriquées pour un but communicatif précis : un instrument de musique ou un stylo. Lorsqu'un medium augmente en importance sociale, il devient un mode et développe une grammaire.[8]

Stöckl (2014) propose également un modèle intégré composé de *mode* et de *médium*. Le *medium* est limité au matériel et à la technologie utilisée pour produire, transmettre et stocker des signes : l'« iPad » ou « la visioconférence ». Un mode existe dans un médium. Un mode est une modalité pour créer des signes, comme les *gestes* ou le *langage oral*. L'avantage de ce modèle réside dans le fait que le *mode* et le *médium* sont séparés. Le *langage* est un mode qui a plusieurs *médialités*[9] : l'oral, l'écrit, le geste.

Norris (2014) ne distingue pas entre le *mode* et le *médium* car un médium ne peut exister sans son usage social.

Selon nous, il est important d'opérer une distinction entre *medium* et *mode* car en visioconférence, les outils techniques sont créés par un designer et ne changent pas au cours de l'interaction. Les modes, quant à eux, sont fluides et émergent de l'interaction. Une distinction entre *mode* et *medium* nous semble nécessaire afin de distinguer entre les outils et la façon dont ils sont employés en interaction.

[8] Une grammaire est un système de ressources qui est partagé socialement. La manière dont ces modes (qu'ils soient visuels, linguistiques, ou autres) sont utilisés est socialement reconnue (Baldry & Thibault, 2006).

[9] « Mediality » (Stöckl, 2014 : 276).

Les propos de Kress & Van Leeuwen (2001) et de Stöckl (2014) renvoient à la distinction qu'opère Bateman (2011) entre *substrat physique*[10] et *mode sémiotique*. Il paraît nécessaire de prendre en compte à la fois le substrat physique d'un mode et ses usages sociaux. Une prise en considération de la matérialité des modes permettra d'opérer cette distinction.

3.3.1.4 Matérialité des modes

Les modes *langage oral* et le *langage écrit* ont certains aspects en commun : lexique, syntaxe et grammaire. Mais ils diffèrent par le son et par la lecture. Le son peut être fort ou faible ; les caractères du texte grands ou petits (Kress, 2014). Il faut aussi prendre en compte le canal de perception sensorielle (visuel, auditif, tactile, olfactif, gustatif, etc.) d'un mode (Stöckl, 2014). Les modes diffèrent donc selon leur matérialité.

Cependant, selon Bateman (2011), il est trompeur de supposer que les modes sont évidents et distincts et qu'ils s'alignent naturellement avec les modalités sensorielles. L'établissement d'un substrat physique est l'un des composants essentiels d'un mode, mais ne suffit pas comme point de départ pour la définition d'un mode. Cela veut dire que les informations visuelles, auditives, tactiles, olfactives et gustatives ne sont pas intrinsèquement distinctes les unes des autres. De ce point de vue, le mode « langage verbal » peut être à la fois sonore et graphique, tout en étant un seul mode sémiotique. La parole orale et le texte écrit peuvent aussi constituer deux modes distincts selon les besoins et les préoccupations du chercheur qui les définit.

Nous faisons une distinction entre *substrat physique* et *mode*. Les substrats physiques existent en soi, tandis que les modes émergent de l'interaction, sont créés par les interactants et sont façonnés par des sociétés. Par exemple, la feuille de papier et le stylo sont des substrats physiques[11], créées par l'homme pour communiquer, tandis que l'écriture est un mode sémiotique par lequel les humains communiquent. La webcam est un outil qui existe en soi, et plusieurs modes sémiotiques passent par ce substrat physique : les gestes, les expressions faciales, les postures. Les substrats physiques sont les canaux de Visu : la webcam, le microphone, le clavier et l'envoi de documents. Les modes sont plus abstraits et doivent être définis.

[10] Un substrat physique (Bateman, 2011) peut être une feuille de papier, un disque vinyle, etc. Pour nous, les substrats physiques sont les canaux de Visu : le microphone, la webcam, et le clavier.

[11] Chanier & Vetter (2006 : 64) emploient le terme « média ».

3.3.1.5 Définir les modes pour la recherche

Les modes ne sont que des abstractions mentales : ils sont séparés les uns des autres de manière analytique, mais non pas de manière constitutive (Baldry & Thibault, 2006). Un mode sémiotique est une unité heuristique qu'il incombe au chercheur de définir en fonction de ses objectifs de recherche (Norris, 2004). Par exemple, on peut dire que plusieurs modes sémiotiques interagissent sur une image d'une magazine : les couleurs, la taille du texte, la distance entre les différents éléments de la page, etc. Le nombre de modes que l'on pourrait définir est sans limites. Selon Norris, si on dit que *disposition* est un mode, celui-ci pourrait comprendre : des meubles, des peintures, des murs, des pièces, des maisons, des rues, etc. Nous pouvons citer d'autres exemples de modes : proxémique, posture, mouvements de tête, gestes, regards, langage oral, texte imprimé, musique, etc.

Les frontières entre les modes sont floues (Kress & Van Leeuwen, 2001 ; Norris, 2004). C'est pour cette raison que Kress & Van Leeuwen (2001) ne se demandent pas quels modes ou quelles combinaisons de modes sont utilisés, mais plutôt quels ensembles de ressources sont produits. Cependant, pour la recherche, les modes doivent être distingués afin d'être analysés. On peut mesurer et classer les modes par matérialité selon qu'ils sont visuels, matériels, durables, ou autres. Le langage oral n'est ni visible ni durable à moins qu'il ne soit enregistré. Les gestes sont visibles mais non durables, alors que le texte imprimé est en principe visible et durable (Norris, 2004). On peut aussi classer les modes par fonction. Chaque mode a une gamme de capacités différentes, et c'est pour cette raison que Kress & Van Leeuwen (2001) soulèvent les questions suivantes : que peut faire un mode de bien ou de mauvais ? Quels sont les signes qui peuvent être produits par ce mode ? Que permet de faire un mode ? Qu'est-ce qu'il est impossible de faire ? Quand on veut représenter quelque chose, quels engagements sont obligatoires avec un mode donné ? Par exemple, dans un livre de biologie, si on décrit une cellule textuellement, on n'est pas obligé de décrire la taille ou la position du noyau. Si, par contre, on dessine une cellule au tableau, on est obligé de s'exprimer sur la taille et la position du noyau.

3.3.2 Les canaux de Visu

Les *substrats physiques* décrits plus haut sont donc les canaux de Visu (cf. chapitre 1). Ceux-ci sont relativement faciles à décrire car ils existent en soi, contrairement aux modes qui ne sont que des abstractions mentales (Baldry & Thibault, 2006) et qui sont définis en fonction des besoins de l'analyste (Norris, 2004). Blin (2016) explique que dans une

perspective cognitiviste, un outil est envisagée, conçue et réalisée par un concepteur. Le chercheur se focalise sur les fonctionnements de base et sur ce qui est physiquement permis par l'outil. Visu est une plateforme de visioconférence qui a été conçue par l'Université de Lyon pour fonctionner avec des ordinateurs. Les affordances de ce dispositif existent grâce à plusieurs concepteurs. Les canaux de Visu sont : le microphone, la webcam, le clavardage et l'envoi de documents.

3.3.3 Les modes sémiotiques

Dans cette section, nous allons décrire les modes sémiotiques et leurs potentiels communicatifs. Nous disons avec Blin (2016) que l'analyse de la CMO ne peut pas être réduite aux dimensions technologiques listées ci-haut (les canaux de Visu). Pour définir les modes, il faut regarder ce qui est perceptuellement saillant (Baldry & Thibault, 2006) tout en gardant en tête les objectifs de l'analyse (Norris, 2004).

3.3.3.1 Parole orale et parole écrite

Puisque les interactants portent un casque audio, tous les sons produits par eux sont perceptibles. Notamment, la langue étrangère est véhiculée par le microphone, et c'est ce qui nous intéresse pour l'enseignement du FLE. Certains phénomènes comme l'intonation, la saillance vocalique et les poses seront annotés, car ils servent à accentuer des éléments que les apprentis-enseignants considèrent importants, comme des items lexicaux.

La parole écrite est véhiculée par le clavardage, et pour nous la parole écrite est un seul mode. Bien que le clavardage puisse être multimodal « en ce sens que son décodage sémiotique fait intervenir des modalités non seulement linguistiques, mais également visuelles comme les binettes[12], le choix de la casse ou de la couleur, la disposition sur la page ou encore la longueur des fils de conversation » (Develotte et al., 2008 : 19), nous ne disposons que des messages monomodaux qui ont été envoyés par la fenêtre de clavardage de Visu.

Nous séparons ces deux modes pour deux raisons. Un des objectifs de notre recherche étant de savoir comment les participants utilisent les canaux de Visu, il semble nécessaire de séparer la parole orale, qui passe par le microphone, et la parole écrite, qui passe par le clavardage. Puis, au niveau cognitif, on suppose que le cerveau humain dispose d'un processeur visuel et d'un processeur auditif (Mayer, 2001). Ces deux processeurs fonctionnent indépendamment l'un de l'autre. Ce phénomène sera revisité plus bas (section 3.3.3.5 en page 99).

[12] Émoticône, ou « smiley » en anglais.

3.3.3.2 Gestes et expressions faciales

Sur l'écran de l'ordinateur, le champ[13] est entouré par une bordure qui sert d'un *méta-commentaire* (Baldry & Thibault, 2006). Ce méta-commentaire indique que tout ce qui est à l'intérieur de la zone est important et sert à créer du sens. A l'intérieur du champ, beaucoup d'éléments sont visibles : l'interlocuteur en face, la salle derrière, et de tierces personnes qui entrent et sortent du champ de la webcam (cf. chapitre 1 figures 5 et 6 en page 29). Si l'interlocuteur se positionne devant la webcam de façon à ce que la tête et le buste soient visibles, on voit les habits, les expressions faciales, les mouvements des bras, les mouvements corporels, la posture, le maquillage, la couleur des cheveux, et encore d'autres choses. Si on tentait d'annoter tout ce qui est visible, l'annotation ne se terminerait jamais. Il faut annoter uniquement ce qui est pertinent pour l'analyse. Notamment, le champ rend visible ce qui est souvent appelé le langage non verbal. La communication non verbale a été définie par (Tellier, 2014a : 103) comme :

> « la façon dont le corps participe de manière plus ou moins directe à la communication. Ainsi, les signaux émis par le corps de notre interlocuteur nous permettent de décoder des intentions, des émotions et des sentiments qui la parole ne dit pas forcément. »

L'importance du « non-verbal » lors des séquences d'explication a été évoquée par De Gaulmyn (1991 : 292) :

> « Il faudrait y ajouter le non-verbal. Parmi les indices non-verbaux, les gestes manuels ont une part importante, les locuteurs montrent les objets, les manipulent quand ils sont sous leurs yeux, ils figurent, miment, représentent spatialement les objets et les activités invisibles. »

Cette citation montre que la multimodalité est considérée comme nécessaire lors de l'interaction didactique. Or pour beaucoup d'auteurs, c'est encore considéré comme secondaire. En effet, la citation ci-dessus a été reléguée à une note de bas de page ! Loin de faire fi des aspects multimodaux de la communication, notre thèse s'inscrit dans un courant de recherche qui a pour vocation de renouveler le logiciel de l'analyse des séquences métalangagières en valorisant l'apport de chaque mode sémiotique.

Comme (Norris, 2004), nous sommes opposé au terme « non-verbal ». Ce terme est inapproprié à une approche multimodale car son usage sous-tend que les indices visuels susmentionnés ne jouent qu'un rôle subordonné à la parole. Selon l'approche multimodale que nous adoptons, aucun mode n'est *a priori* subordonné en permanence. Cependant, il faut prendre en compte la matérialité d'un mode. Pour cette raison, nous parlerons désormais de *ressources visuelles* qui sont rendues visibles et disponibles grâce à la webcam (les gestes, les

[13] Ce qui est filmé par la webcam.

expressions faciales, les hochements de tête, et les postures). Nous allons décrire les modes visuels qui seront pris en compte lors de nos analyses.

Comme la citation ci-dessus met en évidence, les ressources visuelles englobent plus que les gestes manuels. Les études gestuelles ont connu un renouvellement d'intérêt au 20e siècle. Kendon (2004), en faisant un survol des grands noms du domaine, fait remarquer qu'il existe toute une gamme de ce qui peut être considéré comme faisant partie du mode *gestes*, ou du *médium gestuel*. Alors que McNeill, Wundt et Efron (cités par Kendon, 2004) se concentrent uniquement sur les bras et les mains, d'autres comme Quintilian et Austin mentionnent le visage et d'autres parties du corps. Kendon (2004 : 107) explique :

> « Les humains ont à leur disposition le *médium gestuel* qui peut être utilisé de différents moyens, et à partir duquel peuvent être façonnées plusieurs formes d'expression. Les formes qui sont façonnées dépendent des circonstances d'usage, des objectifs communicatifs auxquels elles sont destinées, et de la manière dont elles sont censées être utilisées par rapport aux autres modes d'expression disponibles » (notre traduction)[14].

Nous allons analyser les gestes et mimiques faciales ensemble sous l'onglet « ressources visuelles ». Nous donnons priorité au canal ici, et non pas à une séparation des parties du corps, à cause de la particularité de nos données. En effet, les échanges que nous observons se déroulent en visioconférence, ce qui implique l'usage des canaux de *Visu*. Ce qui nous intéresse est la façon dont les apprentis-enseignants s'emparent des outils à leur disposition. Les usages de gestes et d'expressions faciales font partie tous les deux de l'usage de la webcam. En raison de la taille limitée du champ de la webcam, les mimiques faciales et les gestes sont très proches. A cet effet, Cosnier & Develotte (2011) ont remarqué que les mimiques faciales jouaient un rôle compensatoire dans leur corpus d'interactions en visioconférence à cause de la difficulté qu'avaient les interlocuteurs à rendre visible leurs gestes manuels.[15]

Les effets des expressions faciales sur la compréhension ont été explorés par Sueyoshi & Hardison (2005). Ces auteurs se sont attachés à mesurer les effets des indices visuels sur la compréhension en langue étrangère d'une vidéo d'un locuteur natif en train de donner un cours de céramique pour débutants. Ils ont réparti quarante-deux apprenants d'anglais dans une université américaine de niveaux intermédiaire et avancé en trois groupes. Le premier

[14] « Humans have at their disposal the *gestural medium* which can be used in many different ways and from which many different forms of expression can be fashioned. What forms are fashioned will depend upon the circumstances of use, the communicative purposes for which they are intended, and how they are to be used in relation to the other media of expression that are available » (Kendon, 2004 : 107).

[15] Cela ne veut pas dire pour autant que les gestes et les expressions faciales entretiennent le même lien avec la parole, ou qu'ils relèvent des mêmes processus de production et de perception. Ce positionnement est donc uniquement valable pour nos données où le champ n'offre qu'une image réduite de l'interlocuteur. Afin de faire des propositions didactiques, il faut dans un premier temps comprendre la façon dont la webcam est utilisée.

groupe a regardé l'enregistrement vidéo sans images. Le deuxième a regardé la vidéo où seule la tête était visible, et le troisième a regardé la vidéo où tout le corps de l'enseignant était visible. Après avoir visionné l'enregistrement, les sujets ont passé un test de compréhension orale. Les chercheurs ont trouvé que ceux qui ont vu les images ont compris significativement plus d'informations qui ceux qui n'avaient accès qu'au canal audio de l'enregistrement. De manière intéressante, il n'y avait pas de différence entre le groupe qui a vu le visage de l'enseignant et le groupe qui a vu tout le corps y compris les gestes. Les auteurs concluent que les indices visuels, que ce soit des expressions faciales, les gestes, ou les deux, aident à la compréhension en langue étrangère.

3.3.3.2.1 Les gestes manuels

Les gestes que l'on fait quand on parle, autrement dit les gestes coverbaux ou des gestes spontanés, sont fortement liés au langage et aident à communiquer (Tellier, 2012). Ils aident celui qui parle à organiser ses pensées et à trouver ses mots, et ils aident le destinataire à décoder le sens de l'énoncé (Stam & McCafferty, 2008). Indissociables et complémentaires, les gestes et la parole doivent être pris en compte ensemble pour pouvoir appréhender la totalité du message émis par un interlocuteur qui parle (McNeill, 1985, 1992, 2005). Bien que la parole puisse être comprise sans gestes, au téléphone par exemple, les sujets saisissent plus d'informations lorsqu'ils voient les gestes et entendent la parole en même temps (Beattie & Shovelton, 1999). Kelly, Özyürek, & Maris (2010) ont démontré expérimentalement que si les gestes et la parole d'un interlocuteur sont contradictoires, le destinataire aura beaucoup plus de difficulté à comprendre le message. Ceci rejoint l'avis de McNeill selon lequel les gestes et la parole sont indissociables. L'hypothèse selon laquelle les gestes sont allocentrés, ou faits pour l'autre, a été confirmée par Alibali, Heath, & Myers (2001) qui ont trouvé que les sujets produisent plus de gestes représentationnels quand l'interlocuteur est visible. Ces résultats ont été corroborés par Mol, Krahmer, Maes, & Swerts (2011) qui ont trouvé que les sujets produisent plus de gestes représentationnels en visioconférence quand ils savent qu'ils sont vus par un interlocuteur.

Les gestes intéressent non seulement les linguistes, mais aussi les didacticiens. Dabène, Cicurel, Lauga-Hamid, & Foerster (1993 : 72) remarquent que le non verbal[16] « fait partie intégrante de tout processus de communication qui peut être observé, planifié, vécu ou subi par ceux qui participent à une situation pédagogique donnée ». Elles regrettent que les manuels scolaires n'abordent pas la gestuelle de la culture de la langue enseignée, car les

[16] Nous reprenons ce terme pour rester fidèle au langage des auteures.

gestes sont non seulement linguistiques, mais aussi culturels.[17] Elles remarquent que la proxémique, le regard et le rire sont « trois aspects non-verbaux intrinsèques de la situation pédagogique en classe de langue étrangère » (Dabène et al., 1993 : 74). Les gestes que font les apprenants d'une langue étrangère sont intéressants car ils renseignent sur l'état d'avancement de l'apprentissage (Gullberg, 2008 ; Stam, 2007), mais ce sont les gestes des enseignants qui seront l'objet de focalisation dans le cadre de notre thèse.

Nous nous intéressons aux gestes produits par les enseignants lorsqu'ils enseignent la langue étrangère (Cadet & Tellier, 2007 ; Gadoni & Tellier, 2014 ; Gullberg, 2008 ; Sime, 2006, 2008 ; Tellier, 2005, 2008b, 2008a, 2009, 2012, 2014b ; Tellier, Michel, & Wolff, 2014a, 2014b ; Tellier & Stam, 2010, 2012). Il s'agit d'étudier ce que Tellier, (2008a : 2)[18] nomme les *gestes pédagogiques*. Le geste pédagogique « est principalement un geste des bras et des mains (mais il peut aussi être composé de mimiques faciales) utilisé par l'enseignant de langue dans un but pédagogique. L'objectif premier est de faciliter l'accès au sens en LE ». Les gestes pédagogiques définis par Tellier servent à informer, animer et évaluer. De manière similaire, Sime (2006, 2008) nomme trois fonctions des gestes des enseignants de langue étrangère : les fonctions cognitives, affectives, et organisationnelles. Celles-ci seront discutées dans le prochain chapitre. Nous plaidons avec ces auteurs pour une vue holistique du discours pédagogique de l'enseignant qui comprend non seulement la parole oral et écrite mais aussi les mouvements corporels/kinésiques.

3.3.3.2.1.1 Gestes et apprentissage du lexique

Les gestes pédagogiques sont utiles tout particulièrement pour l'apprentissage du lexique. Lazaraton (2004 : 89), qui a observé la façon dont une professeure d'anglais langue étrangère utilisait ses gestes pour décrire des verbes et des noms, remarque que l'explication lexicale constitue « peut-être la situation prototypique de l'usage langagier qui requiert le déploiement explicite de gestes à des fins pédagogiques » (notre traduction)[19]. Selon cette auteure, le gestuel doit faire partie de toute étude sur l'input langagier, car c'est un aspect saillant et essentiel du répertoire de l'enseignant.

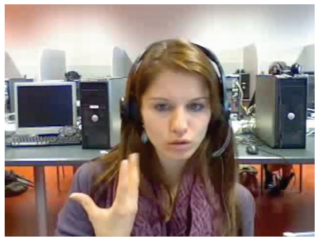
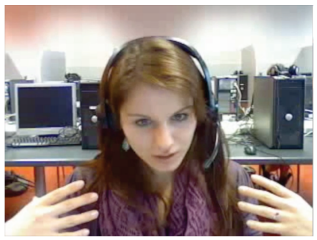
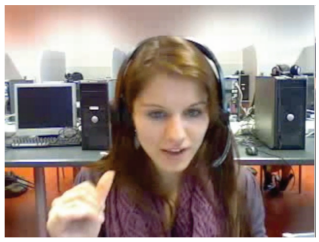
Kelly, McDevitt, & Esch (2009) remarquent que les mots sont arbitraires parce que la plupart du temps, il n'y a aucun lien direct entre la forme phonologique d'un mot et le concept

[17] Cela est en train de changer.

[18] Pagination du manuscrit auteur.

[19] "And one sort of information that is commonly conveyed in the ESL classroom is the meaning of vocabulary words, which is perhaps the prototypical language use situation requiring the explicit deployment of gestures for teaching purposes" (Lazaraton, 2004: 89).

qu'il désigne. Par exemple le mot « boire » : on remarque qu'il n'existe aucune relation entre la forme phonologique du mot et l'action qu'il désigne. En revanche, il y a une relation entre un concept et un geste coverbal qui le désigne, surtout si le geste a un haut degré d'iconicité. Par exemple, le geste qui consiste à imiter la préhension d'un verre et le porter à sa bouche, a une grande ressemblance à l'acte physique de boire un verre. Dans l'exemple suivant, Mélissa représente par des gestes trois items lexicaux : *boissons*, *chaud*, et *boire* quand elle essaie de comprendre le discours d'Alejandra.

Parole orale	**Image de Mélissa**	**Phénomènes observés**
1 ALJ eh: "in" euh "spain"\ "i-" hum: (0,9) beaucoup de choses/ (.) euh (inaud.) euh "hot"/ (0,3) [ɛs] (1,0) 2 MEL °ch- d-° oui/ **de boisson chaude/ (0,7)**		Geste iconique représentant une boisson
3 ALJ (oui) (.) xx (0,5) oui "and:" (0,4) euh d- eh euh (0,6) euh le "beer"/ [euh] 4 MEL [ah] (0,7) ah d'accord il faut très chaud en espagne (1,3) il [faut **chaud en es]pagne/** 5 ALJ **[euh:] (0,9) oui parce que je** habite "in: (.) canary island\" (1,4) l-		Geste emblématique représentant la chaleur
6 MEL et alors vous (0,2) **vous buvez des bières/ (.) pour vous ra[fraichir** un peu\] 7 ALJ [non. n-] normalement non mais ((rires)) (0,2) 8 MEL d'accord\		Geste emblématique représentant l'acte de boire

Exemple 3 : Des gestes pour faciliter la compréhension

Dans cet exemple, Alejandra, d'origine espagnole, répond à une question posée par son binôme concernant la consommation de boissons alcoolisées au travail en Espagne. Mélissa essaie de la comprendre et de reformuler, en faisant des gestes iconiques et emblématiques afin de faciliter la compréhension.

Les gestes nous intéressent donc pour leur potentiel en tant qu'outil pédagogique parce qu'ils font le lien entre des mots connus et des mots nouveaux. Des études empiriques ainsi que des études en neurosciences ont démontré que l'apprentissage de mots est intrinsèquement lié au corps (Macedonia, 2014). Les gestes ont donc des effets profonds et

non négligeables sur la compréhension et la mémorisation du lexique en langue étrangère. Le geste comme outil d'apprentissage du lexique a fait l'objet de plusieurs études.

Allen (1995) a mesuré l'effet qu'avaient les gestes emblématiques[20] sur l'apprentissage d'expressions françaises par des apprenants américains en première année d'études universitaires. En se basant sur l'hypothèse de Craik & Tulving (1975) selon laquelle le traitement profond d'un item langagier crée des traces mémorielles durables, elle a postulé que les gestes emblématiques aideraient au rappel d'expressions françaises. Elle a trouvé que les étudiants qui ont appris les expressions accompagnées de gestes emblématiques se sont rappelés de significativement plus d'expressions que ceux qui les ont apprises sans gestes. Les résultats de cette étude ont amené Allen à plaider pour que les enseignants de langues étrangères utilisent des gestes pour élaborer l'apprentissage dans la mesure où leur emploi est faisable et naturel. Ces résultats ont été corroborés par Macedonia, Müller, & Friederici (2011) qui ont trouvé que la production de gestes iconiques par des apprenants a un impact positif sur la mémorisation de mots.

Tellier (2005) a étudié l'impact de gestes et d'images sur la mémorisation à court terme de mots français dans une classe de grande section de maternelle en France. Elle a fait visionner à deux groupes d'enfants des vidéos d'un sujet en train de lire une suite de mots sans bouger le corps, avec simplement des gestes, ou des images. Elle a trouvé que les enfants qui avaient entendu les mots accompagnés d'images ou de gestes ont retenu plus de mots. Elle a donc conclu que le support visuel, que ce soit des gestes ou des images, aident à la mémorisation à court terme. Dans une autre étude, Tellier (2008b) a démontré que les enfants français de quatre à cinq ans qui apprenaient des mots anglais en reproduisant un geste les mémorisaient mieux que des enfants qui apprenaient les mêmes mots en voyant une image seulement. Tout porte à croire que l'encodage kinésique renforce l'apprentissage et la mémorisation de mots. Allaf-Akbary & Mosalli (2015) se sont intéressés aussi à l'efficacité des gestes et des images pour l'apprentissage de mots en anglais langue étrangère. En menant une expérience auprès d'enfants iraniens apprenant l'anglais, ils ont trouvé que les enfants qui ont appris des mots en produisant un geste et en voyant une photo ont appris significativement plus de mots que des enfants ayant vu une photo uniquement. Les auteurs concluent que les enseignants devraient enseigner du vocabulaire avec des gestes et veiller à ce que les apprenants répètent les gestes.

[20] Un geste emblématique a été défini par (Ekman & Friesen, 1969) comme un geste qui a une traduction orale exacte et qui est connu par tous les membres d'un groupe ou d'une culture. Par exemple, le pouce levé qui aux États-Unis et en France signifie l'accord des parties.

Kelly et al., (2009) ont mené deux expériences, une expérience comportementale et une expérience neuroélectrique[21], auprès d'anglophones adultes en train d'apprendre des mots japonais. Les chercheurs ont mesuré l'impact des gestes sur la mémorisation de mots. Contrairement aux études citées ci-dessus, les sujets ont regardé les gestes mais ne les ont pas reproduits. Malgré cette différence, les gestes ont toujours été utiles à la mémorisation de mots. Les résultats montrent que lorsque le contenu sémantique d'une définition est distribué entre la parole et le geste, le mot est mieux appris et mieux retenu. Les gestes constituent un vrai outil d'explication lexicale : ils ne servent pas uniquement à capter l'attention de l'interlocuteur[22], mais à véhiculer le sens d'un item lexical. Selon cette étude, entendre une définition orale douze fois serait moins efficace qu'entendre une définition six fois accompagnée d'un geste correspondant. Autre avantage : la définition accompagnée d'un geste prend moins de temps que de répéter une définition deux fois.

Pour la seconde expérience, Kelly et al., (2009) ont mesuré deux signaux neuro-électriques lors d'un rappel de mots. Les résultats comportementaux ont corroboré ceux de la première expérience parce que le groupe ayant vu les gestes s'est rappelé significativement plus de mots que le groupe ayant uniquement entendu la parole. Les résultats neuro-électriques ont démontré que les gestes aident à stocker le sens des mots. Les gestes peuvent approfondir les traces mémorielles en les rendant plus imagées et plus incarnées, ce qui renforce la mémorisation.

Les gestes sont utiles non seulement pour la mémorisation d'items lexicaux, mais pour leur explication aussi. Tellier & Stam (2012) ont comparé la façon dont de futurs enseignants de français langue étrangère expliquaient des verbes d'action à des natifs et à des non natifs. Elles ont trouvé que lorsque ces mêmes verbes ont été expliqués à des non-natifs, les gestes que produisaient les apprentis-enseignants duraient plus longtemps, étaient plus illustratifs, et plus amples. On peut supposer donc que les enseignants de langues étrangères considèrent les gestes comme un outil de valeur lors des séquences d'explication lexicale.

Pour résumer, les gestes pédagogiques incluent les gestes manuels et les expressions faciales. Ils véhiculent des informations sémantiques et constituent un outil pédagogique de grande valeur, particulièrement pour l'explication du lexique. Les fonctions précises des gestes pédagogiques, ainsi que leurs annotations, seront abordées dans le chapitre 5.

[21] Une expérience neuroélectrique consiste à brancher sur la tête des sujets des électrodes pour étudier les signaux électriques du cerveau.

[22] Selon Raffler-Engel (1980, cité dans Stam & McCafferty, 2008), les apprenants font plus attention quand l'enseignant fait des gestes iconiques et emblématiques, maintient une posture droite et garde un contact oculaire avec les apprenants.

3.3.3.3 Regards

Les regards dans une salle de classe servent entres autres à gérer les tours de paroles (Azaoui, 2014). Or les regards présentent des particularités sur des écrans. Beattie & Shovelton, (1999) expliquent que les interlocuteurs ont tendance à regarder le visage et les yeux la plupart du temps en face à face, ce qui n'est pas forcément le cas lorsque des sujets regardent des enregistrements vidéo. Le contact oculaire est particulier en visioconférence car si on regarde les yeux de son interlocuteur sur l'écran, celui-ci aura l'impression qu'on ne le regarde pas, car le regard sera détourné de la webcam. Inversement, pour donner l'impression à l'interlocuteur qu'on le regarde dans les yeux ce qui a un effet stimulatoire sur l'auditeur, on doit regarder directement vers la webcam et non pas vers l'écran, ce qui veut dire qu'on ne le regarde pas (De Chanay, 2011). Par conséquent, le contact oculaire est médié et le regard devient une stratégie pédagogique (Satar, 2013). Cette auteure remarque que concernant le regard, il y a toujours une différence entre ce qui est voulu et ce qui est perçu. Certains interlocuteurs s'adaptent à ce paradoxe des regards mentionné ci-dessus se sentent regardés lorsque l'interlocuteur regarde l'écran.

3.3.3.4 Proxémique

La position de la caméra, l'angle et la distance ont tous des effets sur le sens qui est produit par une caméra et sur la communication (Baldry & Thibault, 2006). En visioconférence, on parle plutôt de *distance virtuelle* ou de *distance simulée*. Les apprentis-enseignants choisissent ce qu'ils veulent montrer, ce qui a amené Guichon (2013 : 107) à parler de « différents degrés d'investissement de la webcam ». Selon ce barème, un degré faible ou nul d'investissement serait un apprenti-enseignant qui ne montre pas la totalité de son visage devant la webcam. Un haut degré d'investissement serait un apprenti-enseignant qui se montre expressif et qui regarde directement vers la webcam de temps en temps.

Pour résumer, plusieurs modes sémiotiques passent par la webcam, notamment les gestes, les expressions faciales, et les regards. Les participants disposent d'un contre-champ qui leur permet de regarder leur propre image et d'ajuster ainsi l'image qu'ils projettent sur l'écran. Satar (2013) a trouvé que le contre-champ peut être utile ou déstabilisant. Maîtriser son image devant la webcam est une compétence pédagogique à développer pour les enseignants en ligne. Nous ne pouvons pas dire avec certitude quels gestes ou expressions faciales sont produits avec une intention communicative, mais nous allons voir que le concept

de *recepient design*[23] (Sacks, Schegloff, & Jefferson, 1974) semble être valable aussi en visioconférence. Nous sommes d'ailleurs du même avis que Kendon (2004) qui soutient que l'homme a une capacité innée à distinguer les mouvements communicatifs. L'auteur formule son hypothèse en s'appuyant sur des études qui ont démontré que les groupes de personnes ayant visionné quelqu'un en train de parler peuvent dire avec quasi unanimité quels mouvements corporels sont produits avec une intention communicative et conscientisée. Nous espérons que cette capacité est transmise au chercheur qui annote les enregistrements.

3.3.3.5 Les images et vidéos

Visu permet aux utilisateurs d'envoyer des images à condition qu'elles soient enregistrées sur la plateforme en amont de la séance. Nous verrons dans le chapitre 7 que certains apprentis-enseignants envoient des photos lors des séquences d'explication lexicale. Les images offrent un accès direct à une représentation mentale de l'item lexical et permettent un double encodage, ce qui facilite la mémorisation et le rappel (Chun & Plass, 1997). Selon la théorie du *double encodage*, les traces mémorielles crées par l'apprentissage d'un item lexical sont susceptibles d'être plus profondes si elles sont à la fois verbales et visuelles (Clark & Paivio, 1991 ; Paivio, 2006).

Selon le *principe de multimédia* (Fletcher & Tobias, 2005 ; Low & Sweller, 2005 ; Mayer, 2001), les apprenants apprennent mieux quand les mots et les images sont combinés. Mayer (2001) a démontré expérimentalement que quand on ajoute des images à des textes d'explication scientifique, l'apprentissage est plus profond. Même si les images et les mots véhiculent les mêmes informations, ils diffèrent cognitivement car les mots utilisent le processeur verbal du cerveau humain tandis que les images utilisent le processeur visuel. Cependant, deux conditions sont nécessaires pour que les images et les mots se complètent harmonieusement. Il faut d'abord que les informations véhiculées par chaque mode ne soient pas redondantes, c'est-à-dire que la présence du texte et de l'image doit être nécessaire afin de comprendre le message. Ensuite, il faut que le processeur visuel ne soit pas surchargé. Selon cette conception cognitive de la multimodalité, une image peut être accompagnée de commentaires oraux sans surcharger la mémoire de travail, car les informations seront divisées entre les deux processeurs. En revanche, si on envoie une image et on la commente par clavardage, la charge cognitive sera trop élevée car le processeur visuel sera surchargé tandis que le processeur auditif ne sera pas utilisé. La figure ci-dessous donne une

[23] Selon ce concept, un locuteur construit son discours en montrant une orientation et une sensibilisation à ses interlocuteurs.

représentation de cette conception cognitive. Nous ne représentons que la première étape du modèle cognitif où les stimuli audiovisuels entrent dans la mémoire de travail par la mémoire sensorielle avant d'être traités et éventuellement stockés dans la mémoire à long terme.

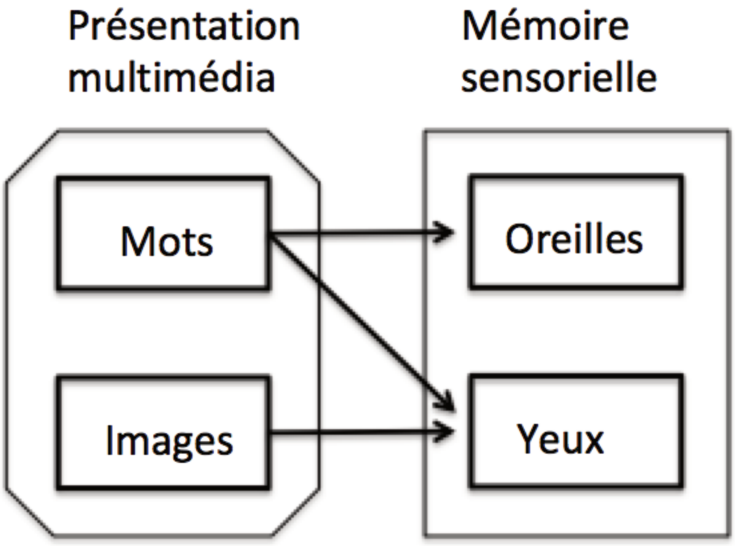


Figure 3 : Début du modèle cognitif de Mayer (2001 : 44)

Plusieurs études expérimentales ont démontré que les images peuvent être bénéfiques pour l'apprentissage du vocabulaire. Lotto & De Groot (1998) ont comparé l'apprentissage de mots étrangers par images et par traduction. Ils ont trouvé que le rappel productif était plus rapide si le mot avait été appris par traduction, et que les rappels étaient plus rapides si le stimulus correspondait à la méthode d'enseignement. Ils concluent que si un enseignant souhaite utiliser une image lors d'une explication lexicale, celle-ci devrait être une valeur ajoutée et non pas une remplaçante d'une explication verbale.

Kost, Foss, & Lenzini (1999) ont comparé l'efficacité de gloses par texte, par images, et par une combinaison de texte et images. Ils ont trouvé que les définitions textuelles seules étaient plus efficaces que les images seules, mais qu'une combinaison de texte et images étaient la méthode d'enseignement la plus efficace. Pour les meilleurs résultats, comme pour l'étude de Lotto & De Groot, le type de rappel doit correspondre au type d'apprentissage (si on apprend un mot par une image, le rappel sera le plus rapide par image). Ces auteurs plaident donc pour que les items lexicaux soient expliqués avec des images car c'est non seulement efficace mais aussi motivant.

Les apprentis-enseignants peuvent aussi envoyer des vidéos tirées de l'Internet. Celles-ci servent de supports pédagogiques pour des exercices de compréhension orale et pour démarrer des discussions sur certains enjeux touchant le monde du commerce en France comme les congés payés, les pauses café et la grève. Il semble que l'usage de vidéos est

bénéfique pour l'acquisition du lexique. Al-Seghayer (2001) a mené une étude auprès de trente apprenants d'anglais pour comparer l'apprentissage de mots par vidéo et texte ensemble, images et texte ensemble, et texte seulement. Il a trouvé que les mots qui ont été appris par vidéo et texte ensemble étaient mieux mémorisés que ceux appris par images et texte ensemble, et que les mots appris par images et texte ensemble étaient mieux mémorisés que ceux appris par texte seulement. Or, dans notre corpus, les vidéos qui sont envoyées par les apprentis-enseignants n'ont pas pour objectif d'expliquer des items lexicaux. Cependant, ils font émerger des lacunes lexicales qui déclenchent par la suite des séquences d'explication lexicale.

Les supports audiovisuels sont très efficaces pour l'apprentissage du vocabulaire quand ils sont combinés avec des définitions verbales comme des traductions, et quand les apprenants peuvent choisir le mode de présentation. Plass, Chun, Mayer, & Leutner (1998) ont trouvé que les apprenants apprennent plus de mots quand ils ont accès à plusieurs modes et quand le mode d'enseignement correspond à la préférence modale de l'apprenant. Ces chercheurs ont étudié l'apprentissage de mots allemands auprès de 103 étudiants américains qui lisaient une courte histoire en allemand sur un écran d'ordinateur. En cliquant sur un mot inconnu, les étudiants avaient le choix entre une traduction seule (orale et écrite), et la combinaison d'une traduction et une animation visuelle (image ou vidéo). Les chercheurs ont trouvé que les mots étaient mieux appris quand une traduction était combinée avec une animation visuelle, ce qui valide leur hypothèse selon laquelle l'apprentissage est plus profond quand les informations sont présentées multimodalement. En outre, les chercheurs ont trouvé que les apprenants ont appris plus de mots quand ils ont sélectionné le mode qui correspondait à leur préférence modale. Grâce à un pré-test, ceux-ci ont catégorisé les étudiants selon leur préférence modale : certains étudiants préféraient des supports visuels tandis que d'autres préféraient des supports verbaux. Les apprenants visuels ont bénéficié de l'ajout d'informations visuelles, ce qui n'a pas été le cas pour les apprenants auditifs. On peut dire que les apprenants devraient avoir un choix du mode d'enseignement.

3.3.4 Synthèse : les ressources sémiotiques

La visioconférence est un médium qui est constitué de plusieurs canaux : la webcam, le microphone, le clavier, et les documents. Plusieurs modes passent par ces canaux : la parole orale, la parole écrite, les expressions faciales, les gestes, et les images. Nous avons sélectionné ces modes pour l'analyse car ils sont susceptibles de servir à conduire des

explications lexicales. Nous avons synthétisé les substrats physiques et les modes dans le schéma ci-dessous.

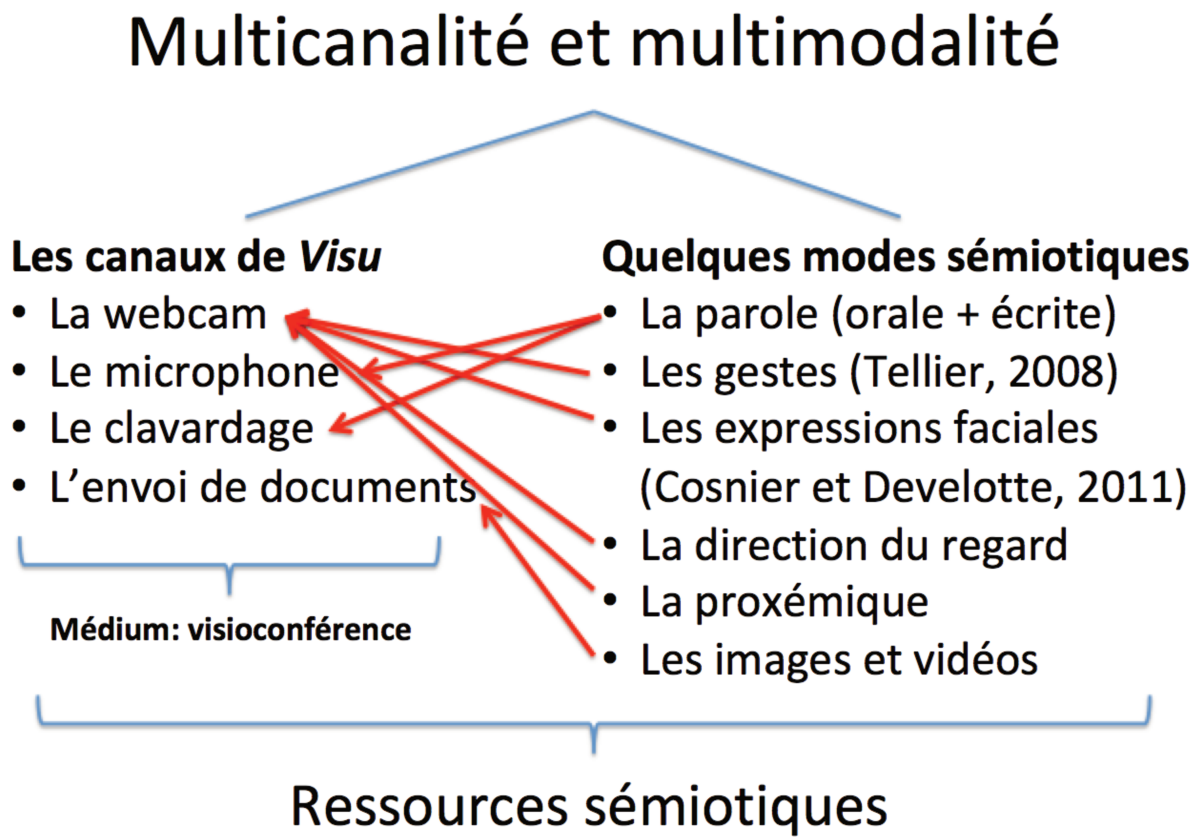


Figure 4 : Les ressources sémiotiques

A gauche de la figure 4, nous avons les substrats physiques, autrement dit les canaux de Visu. L'ensemble de ces ressources matérielles constitue le médium *visioconférence*. A droite, nous avons les modes sémiotiques que nous avons décrits. Ceux-ci passent par les canaux de Visu. L'ensemble de ces ressources physiques et abstraites constitue les *ressources sémiotiques*.

3.3.5 Sélection des modes sémiotiques

Face à un choix de modes, un interlocuteur sélectionnera le mode le plus adapté ou les modes les plus adaptés à faire du sens (Baldry & Thibault, 2006). Par exemple, lors d'une panne technique en visioconférence où le microphone ne fonctionne pas, le clavardage sera le mode le plus adapté à communiquer (Azaoui, 2017). Lors d'une séquence d'explication lexicale, si la combinaison d'un geste et une description orale est la méthode la plus adaptée à faire comprendre le sens de l'item, ces deux modes seront sélectionnés par l'apprenti-enseignant. De même que chaque enseignant a son propre profil gestuel (Tellier et al., 2014), chaque utilisateur de la visioconférence a un profil interactif différent ce qui fait que certains

privilégieront la webcam plus que d'autres (Develotte, Guichon, & Vincent, 2010). Les apprenants ont des profils différents aussi : « chaque individu appréhende différemment les divers supports pédagogiques en fonction de sa modalité (kinesthésique, visuelle, auditive) d'apprentissage privilégiée » (Cadet & Tellier, 2007 : 72). L'étude de Plass et al. (1998) citée ci-dessus soutient l'idée que le profil multimodal de l'apprenant et à prendre en compte. Notre travail consiste à analyser les modes sélectionnés par les apprentis-enseignants pour conduire des séquences d'explication lexicale. Une fois sélectionnés, les modes entrent en interaction les uns avec les autres.

3.3.6 Interaction entre les modes

Bien qu'un mode puisse être utilisé seul[24], les modes sémiotiques sont normalement en interaction permanente les uns avec les autres. Quand plusieurs modes sont sélectionnés, ils créent du sens ensemble. Les uns n'existent pas indépendamment des autres, et l'effet total n'est pas la simple somme de tous les modes. C'est pour cela qu'il faut prendre compte l'interaction entre les modes. Nous avons retenu deux concepts ressortant des études multimodales : celui de l'intégration des modes et celui de la hiérarchisation des modes.

3.3.6.1 Intégration des modes

Un texte multimodal intègre différents modes. Les ressources visuelles et verbales sont combinées pour former des unités syntagmatiques (Van Leeuwen, 2004). Selon le *principe de l'intégration des ressources*, ou intégration des modes, (Baldry & Thibault, 2006), ceux-ci ne sont pas simplement juxtaposés les uns aux autres, ils sont combinés et intégrés pour former un tout qui est complexe et qui ne saurait se réduire à la somme des composants. Les modes ont des effets les uns sur les autres. Ils se co-contextualisent d'une manière qui serait imprévisible si on les considérait comme étant des modes complètement distincts.

3.3.6.1.1 Les gestes et la parole

Prenons l'exemple de l'interaction entre la parole et les gestes, interaction qui a été beaucoup étudiée (Beattie & Shovelton, 1999; Kendon, 2004 ; McNeill, 1985, 1992). Les gestes et la parole forment un ensemble qui ne se réduit pas à la somme des composants. Les gestes et la parole doivent être pris en compte ensemble car aucun des deux modes ne prime *a priori* sur l'autre. La parole s'adapte au geste et vice versa (Kendon, 2004). Par exemple, un interlocuteur peut faire une pause à l'oral pour donner place à un geste. De même, les gestes sont planifiés et intégrés à la parole à quel point que la le geste s'aligne parfaitement à la

[24] On pourrait, par exemple, mener une séquence d'explication uniquement par clavardage.

parole. C'est en écoutant la parole que les gestes font du sens car, sauf certains gestes emblématiques, les gestes seuls ne portent pas de sens.[25] Les gestes peuvent par exemple nous renseigner sur le point de vue d'un interlocuteur qui raconte une histoire. L'énonciateur se positionne-t-il comme un spectateur qui observe l'action d'un sujet, ou bien comme le personnage dont il décrit l'action ? Aussi, les gestes sont particulièrement bien adaptés à décrire la taille et la forme des objets. Kendon (2004) liste quatre fonctions du geste concernant la relation qu'il entretient avec la parole :

1. Une fonction modale par laquelle le geste modifie le sens qui est communiqué par la parole (par exemple le geste nous renseigne sur la taille d'un objet qui est décrit oralement) ;
2. Une fonction performative par laquelle le geste définit le type d'acte de parole ou action interactive (par exemple, le geste communique le fait que l'on fait une demande ou offre des informations) ;
3. Une fonction analytique par laquelle le geste ponctue le discours et marque l'organisation et l'importance des composants ;
4. Une fonction interpersonnelle par laquelle le geste indique à qui on s'adresse.

En plus de ces quatre fonctions générales, Kendon ajoute six autres moyens par lesquels les gestes peuvent être complémentaires ou redondants à la parole orale. Même les gestes redondants ne sont pas sans effet sur la communication : ils sont utiles si on n'est pas sûr que l'interlocuteur va comprendre, et ils ont un impact sur la façon dont un énoncé est reçu et expérimenté. Nous avons cité plus haut l'étude de Kelly et al. (2009) qui a démontré que les gestes redondants aident à la mémorisation du lexique et sont rentables temporellement. Ces fonctions montrent à quel point les gestes et la parole s'imbriquent. L'interaction entre le geste et la parole n'est qu'un exemple d'interaction entre modes. En visioconférence, tous les modes sélectionnés par les interactants sont en constante interaction les uns avec les autres. Nous verrons dans l'exemple 5 ci-dessous que le clavardage et la parole peuvent aussi entrer en interaction l'un avec l'autre.

3.3.6.2 Hiérarchisation des modes

Norris, (2004, 2014) a élaboré un cadre théorique pour rendre compte de la *configuration modale* qui fait référence à la relation hiérarchique entre les modes lors de la production du sens. Bien que chaque mode contribue à la production du sens, il existe des

[25] Les langues des signes portent du sens sans parole mais ne font pas partie des gestes coverbaux que nous étudions.

relations de hiérarchie entre les modes. A un moment donné, certains modes sont plus producteurs de sens que d'autres, voire plus importants. Pour reprendre l'exemple du geste et de la parole, si la parole s'arrête à un moment pour faire place à un geste, le geste prime sur la parole lors de la pause et devient le mode le plus important. Norris parle alors de *densité modale* qui peut être atteint soit par intensité modale, soir par complexité modale, soit par les deux à la fois.

Un mode devient très *intense* lorsqu'il occupe une place primaire dans une interaction. Un mode est très intense si :

- D'autres modes sont structurés par ce mode ;
- L'action réalisée par l'acteur à l'aide de ce mode serait affectée et/ou rendue impossible si ce mode n'était pas très intense ou cessait d'exister ;
- Tous les modes incarnés[26] sont focalisés sur ce mode.

Il y a *complexité multimodale* lorsque les modes sont interconnectés[27] de manière complexe.

Prenons le mode *clavardage* afin d'analyser sa densité modale lors de quatre exemples. Le clavardage a une densité modale décroissante dans les quatre exemples suivants.

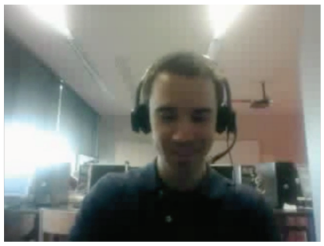
Transcription de la parole	**Image du participant**	**Phénomènes observés**
1 SEV <((clavardage)) **tu es connectée**>		Regard vers le clavier

Exemple 4 : Intensité modale forte

Le clavardage est très intense ici car c'est le seul mode par lequel les interlocutrices communiquent. Lors d'une panne technique, Séverine n'entend pas son apprenante et lui demande par clavardage si elle est connectée. Puisque la communication serait impossible sans le clavardage, ce mode est considéré comme intense.

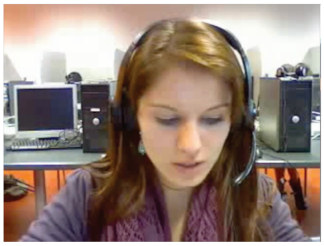
[26] Norris (2004) distingue modes incarnés et modes désincarnés (*embodied modes* et *disembodied modes*). Les premiers englobent toutes les actions qui sont réalisées par le corps humain : regards, gestes, posture, parole orale. Les seconds englobent par exemple des textes, meubles, photos, bâtiments. L'unité d'analyse pour les deux est l'action : les modes incarnés actualisent une action en temps réel, tandis que les modes désincarnés communiquent des « actions figées » qui ont été réalisées auparavant.

[27] Tous les modes qu'utilise un participant sont interconnectés, ce qui fait que l'interaction est susceptible de changer si un mode est ajouté, interrompu/abandonné, ou utilisé différemment. Ce phénomène s'observe souvent lorsque le canal audio cesse de fonctionner.

Transcription de la parole	Image du participant	Phénomènes observés
1 VIC c'est **(0,3) hum:** (2,9) <((clavardage)) Un bobo.> et en fait/ (.) c'est^une contraction de deux mots		Regard vers le clavier

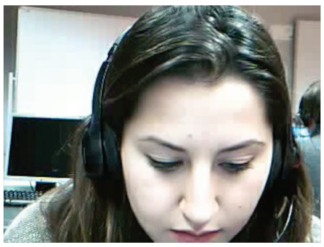
Exemple 5 : Intensité modale forte

Quand Victor donne le mot qui fait l'objet d'une explication lexicale, il choisit de le faire par clavardage. Le clavardage est le complément de sa voix, car sans clavardage la phrase serait incomplète. Voici un bel exemple de ce que Sindoni (2013) appelle *mode switching*, ou l'alternance entre les modes. Kendon (2004), en parlant de l'interaction entre les gestes et la parole, démontre que la parole laisse souvent place au geste via une pause. Dans cet exemple, la parole laisse place au clavardage via une pause de 2,9 secondes.

Transcription de la parole	Image du participant	Phénomènes observés
1 MEL °je l'écris en fait° (0,4) 2 ALJ [les salari]és/ 3 MEL [c'est le mot/] **(0,6) salarié** <((clavardage)) salarié>		Regard vers le clavier

Exemple 6 : Intensité modale moyenne

Mélissa donne un mot-clé de l'activité, mot qui fait aussi l'objet d'une séquence d'explication lexicale. Elle le donne oralement et par écrit, mais contrairement à l'exemple précédent, le clavardage ici est redondant à la voix. Cependant, il est co-contextualisé par un commentaire méta-sémiotique (ligne 1).

Transcription de la parole	Image du participant	Phénomènes observés
1 SAM et vous allez construire ce qu'**on appelle un pense-bête\ (1,6) vous savez c**`que c'est un pense-bête/ <((clavardage)) un pense bête>		Regard vers le clavier

Exemple 7 : Intensité modale faible

Samia introduit le mot-clé de l'activité et de la séquence d'explication lexicale, puis le répète par clavardage. Le clavardage est redondant à la voix et n'est pas co-contextualisé par ce mode. Le clavardage a donc ici une intensité faible. Dans tous les quatre exemples cités ci-dessus, le regard des apprentis-enseignants (mode incarné) est focalisé sur le clavardage, ce qui augmente l'intensité modale du clavardage.

Norris suppose que plus la densité d'un mode est élevée, plus l'acteur prête attention à l'action qui est rendue possible par ce mode. La hiérarchie n'est pas figée ou préétablie et aucun mode ne prime sur un autre en permanence. Pour Norris, si deux modes sont utilisés ensemble, et si on ne peut pas avoir l'un sans l'autre, il y a un *agrégat*. Un agrégat de modes s'observe dans l'exemple 5 ci-dessus où la voix et le clavardage se complètent et on ne peut avoir l'un sans l'autre.

3.3.6.3 Phases multimodales

Nous avons évoqué ci-dessus en page 83 la division des séquences d'explication lexicale en trois phrases : ouverture, noyau et clôture. Baldry & Thibault (2006) ont élaboré la notion de *phase multimodale*. Une phase multimodale est caractérisée par une cohérence entre les fonctions des modes sémiotiques. Lors d'une phase, tous les modes sémiotiques produisent des sens complémentaires afin de réaliser une action. Les phases servent à rythmer le texte, celles-ci pouvant être définies différemment selon la granularité d'analyse. Une phase peut être constituée d'une activité entière, ou bien uniquement de l'ouverture de celle-ci. Les phases d'un texte multimodal sont facilement perceptibles par le lecteur, ce qui veut dire que les points de transition sont aussi perceptibles.[28]

Nous nous proposons d'appliquer la notion de *phase multimodale* au modèle canonique des séquences d'explication lexicale évoqué plus haut. Lors de la phase d'ouverture d'une séquence d'explication lexicale, l'item lexical est repéré, décontextualisé et problématisé. Les modes sémiotiques sélectionnés travaillent ensemble pour réaliser cette action. Lors du noyau, les modes sémiotiques déployés par l'apprenti-enseignant travaillent ensemble pour expliquer une facette de la connaissance de l'item lexical ou pour trouver et donner le *mot juste* (Grossmann, 2011). La notion de phase multimodale de Baldry et Thibault nous permet d'analyser l'apport de chaque mode lors de l'exécution d'une action pédagogique donnée.

[28] Cependant, les frontières entre les phases peuvent parfois être floues et difficilement perceptibles.

3.4 Positionnement épistémologique

Jewitt (2014b, 2014c) liste trois approches de la multimodalité : (1) l'approche sémiotique sociale, (2) l'analyse du discours multimodal[29], et (3) l'approche par l'analyse multimodale interactionnelle. Pour la première approche, on s'intéresse à la façon dont on crée des signes dans un contexte social donné. Le contexte a une influence sur les ressources qui sont disponibles et sur la façon dont celles-ci sont sélectionnées et conçues. On se focalise sur des créateurs de signes. Cette approche est adoptée notamment par Kress & Van Leeuwen. Pour la deuxième approche, on s'intéresse au système et non pas aux créateurs de signes humains. On s'intéresse notamment aux métafonctions de Halliday (1978, cité par Jewitt, 2014b) et on se pose les questions suivantes : que peut-on faire avec un mode ? Que fait-on, compte tenu des choix disponibles ? Bref, on s'intéresse à l'action (Jones, 2014). Pour la troisième approche, on s'intéresse à l'interaction et aux actions qui sont réalisées par les acteurs, car les modes n'existent pas en dehors de l'interaction.

Nous proposons une conception hybride de la multimodalité qui prend en compte, à la fois le support physique (les canaux de Visu) et les modes sémiotiques sélectionnés et employés par les apprentis-enseignants. On se demande avec Demaizière (2000 : 306) : « Comment trouver un chemin raisonnable entre une fascination techniciste et un faux humanisme rétrograde tout aussi critiquable ? » On ne peut pas négliger les modes de Visu car ils ont été construits par un concepteur, et on ne peut non plus ignorer les acteurs humains et succomber à ce que Albero & Thibault (2009 : 57) nomme « la survalorisation de l'innovation technologique » en sciences humaines et sociales, qui a comme conséquences « idéologie dans les positionnements, manque de légitimité épistémique, manque de scientificité des démarches ». En effet, la parole orale est fortement multimodale : intonation, débit, prosodie, gestes, et mimiques faciales constituent autant de phénomènes qui s'imbriquent lorsque les humains parlent, que la communication soit médiée ou non.

Nous considérons que nous étudions d'abord des acteurs humains qui sont en train de réaliser une action. Cette action est médiée par un support technologique, mais les acteurs humains et leurs actions priment sur le dispositif technique et sont au centre de nos analyses. Nous adoptons donc ce que Rabardel (1995) nomme une perspective anthropocentrique. Cette perspective s'oppose à une perspective technocentrique « où l'homme occupe une position résiduelle et où, son activité réelle n'ayant plus de statut propre, elle ne peut, le plus souvent,

[29] Systemic functional multimodal discourse analysis (SF-MDA), termes de O'Halloran, cités par Jewitt (2014b).

être pensée que dans les termes mêmes du processus technique » (Rabardel, 1995 : 20). Selon une approche anthropocentrique, « ce sont les systèmes techniques, les machines qui sont pensées en référence aux hommes et non l'inverse ; où la place de l'homme est première et celle de la technologie définie par rapport à celle-ci » (Rabardel, 1995 ; 24). Cet auteur note tout de même qu'aucune de ces deux perspectives n'est suffisante toute seule. C'est pourquoi nous n'ignorons pas les effets de médiation entretenus par le dispositif.

3.4.1 Intérêt d'étudier les apprentis-enseignants en formation

Les interactions que nous analyserons sont dignes d'intérêt pour des raisons humaines et techniques. Concernant les acteurs humains (apprentis-enseignants), ceux-ci n'avaient jamais donné des cours en visioconférence. L'enseignement des langues étrangères en visioconférence était relativement nouveau pour eux et des pratiques pédagogiques et techniques étaient toujours en développement lors de l'enregistrement des interactions.

Selon Kress & Van Leeuwen (2001), l'importance d'un mode comme les gestes ne vient pas de son potentiel ou de sa valeur, mais du degré dans lequel il a été développé et exploité par une société. Si les gestes occupent une place secondaire pour certains chercheurs travaillant dans le domaine de l'analyse conversationnelle, ce n'est pas à cause des limitations de la matérialité de la ressource, mais parce que la communauté n'a pas investi, n'a pas développé, et n'a probablement pas travaillé cette ressource. C'est pour cela que certains chercheurs comme McNeill (1992) et ses disciples plaident pour une meilleure prise en compte de la gestuelle et de l'interaction entre les gestes et la parole.

Si on déménage d'un appartement de deux pièces pour aménager dans un appartement de cinq pièces, on n'aura pas nécessairement de quoi meubler immédiatement les cinq pièces, faute de meubles et d'habitudes. Quelques mois plus tard, on peut s'attendre à ce que toutes les pièces soient meublées. De même, quand on passe de l'audioconférence à la visioconférence, les acteurs ne sauront pas immédiatement comment tirer profit de la webcam, en utilisant des gestes et mimiques faciales par exemple. Voilà peut-être la raison pour laquelle Guichon & Cohen (2014) n'ont pas trouvé de différences significatives entre l'interaction par audioconférence et l'interaction par visioconférence : les participants n'avaient pas encore développé des pratiques susceptibles d'exploiter tout le potentiel de la plateforme. Kress (2014) parle alors de la *portée* d'un mode : c'est ce que l'on peut faire ou communiquer avec un mode dans une communauté donnée. Les *portées* des modes ci-dessus

sont toujours en développement, ce qui rend les analyses multimodales que nous mènerons non seulement dignes d'intérêt mais nécessaires.

Quand les supports technologiques changent, de nouveaux usages peuvent se développer, mais cela prend du temps. Baldry & Thibault (2006) donnent l'exemple de l'Internet. Avant l'Internet, il existait bel et bien quantité de pratiques de création du sens : texte écrit, radio, télévision, etc. Avec l'invention et l'évolution de l'Internet, une nouvelle infrastructure technologique s'est progressivement mise en place et cette nouvelle infrastructure a permis à de nouvelles pratiques communicatives de se développer : hypertexte, courriel, blog, etc. De même, des pratiques communicatives et pédagogiques existaient bien avant l'émergence de la visioconférence. Et celle-ci nous donne une nouvelle infrastructure technologique par rapport à la salle de classe traditionnelle, ce qui va permettre aux participants de développer de nouvelles façons de communiquer et de combiner les ressources sémiotiques à leur disposition, d'utiliser les modes existants et aussi d'en créer de nouvelles. Cela renvoie à ce que Develotte et al., (2010) nomment les *compétences sémio-pédagogiques* : la capacité à mener une interaction pédagogique en utilisant les modalités à leur disposition. Ces compétences sont en développement lors des échanges que nous observons.

3.5 Conclusion

Dans ce chapitre, nous avons mis en évidence le fait qu'il existe déjà beaucoup d'études sur les séquences de négociation du sens. Celles-ci ont été modélisées pour la conversation ordinaire en présentiel (Varonis & Gass, 1985) et pour l'interaction pédagogique par clavardage (Smith, 2003). La visioconférence semble être un contexte propice à la négociation du sens (Sarré, 2011). Les séquences d'explication lexicale constituent une catégorie de séquences de négociation du sens et ont fait l'objet d'une modélisation générale à trois phases (De Gaulmyn, 1991 ; Fasel Lauzon, 2008, 2014 ; Gülich, 1990 ; Lüdi, 1991). Chacune de ces études, ainsi que celle de Miecznikowski (2005), traite les ressources sémiotiques autre que la parole orale comme des ressources secondaires, et ne prennent donc pas en compte la dimension multimodale de la parole. Or, la parole est profondément multimodale, comme l'ont montré plusieurs études (Kendon, 2004 ; McNeill, 1985, 1992, 2005). Cette multimodalité est toute particulièrement pertinente dans l'enseignement et l'apprentissage d'une langue étrangère, comme l'ont démontré un grand nombre d'études (Allen, 1995 ; Kelly et al., 2009 ; Tellier, 2005, 2008b ; Tellier & Stam, 2010, 2012).

Multicanal, le dispositif *Visu* rend obligatoire une prise en considération de la dimension multimodale de la communication. Puisque multicanalité n'est pas la même chose que multimodalité (Bateman, 2011), une réflexion sur les différences entre *ressources sémiotiques* et *modes sémiotiques* a été présentée. Nous avons décrit quelques modes sémiotiques qui sont saillants en visioconférence et qui sont susceptibles d'être employés à des fins pédagogiques par les apprentis-enseignants lors des séquences d'explication lexicale. La parole orale, la parole écrite, les gestes, les expressions faciales, les regards, la proxémique, et les images sont des modes sémiotiques qui sont véhiculés par les canaux de Visu : la webcam, le microphone, le clavardage, et l'envoi de documents (cf. Figure 4 en page 102). Ces ressources sont utilisables dans la négociation du sens par visioconférence.

Nous adoptons donc une approche multimodale pour montrer l'apport de chacune des ressources susmentionnées lors des séquences d'explication lexicale. Dans une perspective anthropocentrique, nous étudierons à la fois les pratiques pédagogiques des apprentis-enseignants lorsqu'ils conduisent des séquences d'explication lexicale, et la façon dont le dispositif de visioconférence est utilisé, c'est-à-dire la façon dont les modes sémiotiques offerts par les quatre canaux de *Visu* sont sélectionnés, combinés, hiérarchisés et exploités par les acteurs en situation d'explication lexicale.

Nous pouvons dès à présent mettre à jour les questions de recherche que nous avons abordées à la fin du chapitre précédent :

1. Par quels moyens multimodaux les apprentis-enseignants se focalisent-ils sur la forme des items lexicaux ?
2. Par quels moyens multimodaux les apprentis-enseignants se focalisent-ils sur le sens des items lexicaux ?
3. Par quels moyens multimodaux les apprentis-enseignants impliquent-ils les apprenants dans les séquences d'explication lexicale ?

Ces questions seront explorées en détail dans la troisième partie.

Partie II : Méthodologie

Dans la partie précédente, nous avons situé notre projet de thèse dans un courant de recherche qui suit les évolutions technologiques et vise à étudier la CMO et ses applications pédagogiques. Après avoir abordé l'importance de l'enseignement explicite du vocabulaire, nous avons résumé, à l'aide d'études expérimentales, quelques pratiques susceptibles de favoriser la transmission et l'apprentissage de connaissances lexicales. Enfin, nous avons argumenté en faveur d'une approche multimodale pour l'analyse des séquences d'explication lexicale en visioconférence.

Cette deuxième partie se donne pour objectif de présenter le corpus audiovisuel qui nous permettra d'analyser les pratiques pédagogiques et multimodales mises en place par les apprentis-enseignants lorsqu'ils gèrent des explications lexicale par visioconférence. Cette partie se divise en deux chapitres. Le chapitre 4 expliquera les démarches entreprises par une équipe de chercheurs afin de construire un corpus exploitable à partir d'interactions authentiques. Nous valoriserons le travail de montage audiovisuel et de transcription que nous avons effectué lors de notre première année de thèse. Les enjeux techniques, éthiques et épistémologiques seront abordés. Le chapitre 5 donnera lieu dans un premier temps à une réflexion personnelle sur la relation que nous entretenons avec nos données. Puis, nous décrirons les procédés par lesquels nous avons repéré les séquences d'explication lexicale destinées à l'analyse et notre façon d'annoter les phénomènes multimodaux. Une réflexion sur les multiples transformations des données sera menée.

Chapitre 4 : Construction du corpus

4.1 Introduction

Afin d'étudier l'interaction entre apprenants et apprentis-enseignants en visioconférence, un corpus d'interactions authentiques est nécessaire car « c'est à partir de données empiriques que l'on peut […] saisir quelque chose de l'acte d'enseignement » (Cicurel, 2011 : 8). S'il existe en abondance des interactions authentiques, la construction d'un corpus multimodal d'apprentissage nécessite un grand travail collaboratif dont les grandes étapes incluent le design, le recueil des données, l'organisation du corpus global, la post-recherche[1] et le développement de ressources pour la formation des enseignants (Chanier & Wigham, 2016). Ce chapitre va traiter ces trois premières étapes.

Un corpus n'est pas qu'un ensemble de documents authentiques ; un ensemble de données ne devient un corpus qu'à partir du moment où il y a sélection, structuration et partage (Chanier & Ciekanski, 2010). Par sélection, nous entendons le choix d'inclure dans le corpus ou non des participants en fonction de leurs réponses aux formulaires de consentement, et de cibler stratégiquement les interactions à analyser selon nos questions de recherche. En ce qui concerne la structuration des données, celles-ci doivent être rendues exploitables et analysables par des traitements audiovisuels et par des classements par date et tâches abordées. Enfin, la partageabilité des données occupe une place primordiale pour que d'autres chercheurs possédant différentes questions de recherche puissent travailler sur les mêmes données et pour que nos analyses puissent être vérifiées ou complétées.

Ce chapitre se donne pour objectif de décrire la façon dont notre corpus a été construit en collaboration avec différents membres du projet ISMAEL et du laboratoire ICAR. Dans un premier temps, nous allons décrire le design du corpus qui comprend le dispositif pédagogique et le dispositif technique. Dans un deuxième temps, nous allons décrire la façon dont les données ont été recueillies à partir de traces diverses. Ensuite, nous allons décrire notre manière de créer les vidéos à l'aide de divers logiciels de traitement audiovisuel. Puis, nous allons détailler la transcription du corpus sur le logiciel ELAN. Enfin, nous allons décrire les choix qui ont été faits quant au partage et à la sélection des séances qui sont devenues le corpus ISMAEL.

[1] Pendant l'étape de poste-recherche, on crée un corpus distingué pour répondre à des questions de recherche précises.

Avant de détailler la construction du corpus, il convient de rappeler que cette démarche est un travail d'équipe. L'organigramme ci-dessous présente les membres du projet qui ont aidé à la construction du corpus.



Figure 1 : Les membres du projet ISMAEL

4.2 Design

Chanier et Wigham (2016) rappellent qu'on détermine en premier lieu l'objectif et l'orientation de la recherche, ce qu'on souhaite analyser. Il faut créer un scénario pédagogique qui inclue l'environnement multimodal, les participants, et les activités et rôles pour ceux-ci. Ensuite il faut un protocole de recherche qui rende compte des variables à investiguer, les participants, les enjeux éthiques, les méthodes et procédures pour collecter les données, les rôles des observateurs. Tout ceci doit être expliqué de façon à ce que ceux qui n'ont pas participé au recueil des données puissent comprendre la situation.[2] Le scénario pédagogique et le dispositif technique (Visu) ont été expliqués dans le premier chapitre. Nous allons décrire dans ce chapitre la façon dont des données ont été transformées afin d'être analysées.

[2] Nous n'étions pas présent lors de la création du design ni lors du recueil des données, mais il nous incombe néanmoins d'expliquer les démarches qui ont été effectuées avant notre arrivée.

4.3 Recueil des données

Plusieurs types de données ont été recueillis avant notre arrivé à Lyon : les plans des cours, des captures d'écran dynamiques des interactions, des enregistrements vidéo des interactions par une caméra hors-champ, des post-interviews avec les apprentis-enseignants et les apprenants, et les enregistrements des débriefings. Avant la construction du corpus, nous avons pris en compte des enjeux éthiques qui deviennent de plus en plus importants en sciences humaines et sociales. Ces enjeux éthiques seront discutés à la fin de cette section.

4.3.1 Plans des séances

Tous les plans de séances ont été commentés et validés par Nicolas Guichon et Françoise Blin. Ciara Wigham et Sylvie Thouësny les ont déposés dans un dossier partagé en ligne. Les plans de séances consistent d'activités, mots-clés, images, et objectifs.

4.3.2 Captures d'écran

Comme Rivens Mompean & Guichon (2009), nous travaillons principalement sur des captures d'écran. Lors des séances, un logiciel de capture d'écran enregistrait la totalité des interactions. A chaque fois qu'un participant a allumé sa webcam et s'est connecté, un fichier audiovisuel a été crée et stocké sur un serveur. Ces données audiovisuelles sont disponibles à deux endroits : sur la salon de rétrospection de Visu et dans un dossier partagé. Ces enregistrements constituent la plus grande partie de notre corpus.

4.3.3 Caméras extérieures

Lors des séances, certains interactants lyonnais ont été filmés de l'extérieur afin de rendre visible le hors champ. Ces enregistrements ont été faits par Ciara Wigham et une étudiante de M2DidLang (Amina Dib). Ces vues extérieures ont permis à Guichon et Wigham (2016) de constater que la plupart des gestes faits par les apprentis-enseignants ont été faits hors champ et par conséquent sont invisibles aux apprenants.

4.3.4 Bilans

Après chaque séance, les apprentis-enseignants ont créé sur Visu un bilan pour chaque apprenant. Ces bilans contiennent des commentaires correctifs écrits et oraux ainsi que des appréciations générales. Ces bilans ont été déposés par Sylvie Thouësny dans un dossier partagé.

4.3.5 Débriefings

Le lendemain de chaque séance, les apprentis-enseignants ont discuté de l'interaction passée avec leur formateur, Nicolas Guichon. Ils ont dû sélectionner un moment de l'interaction qui s'était bien passé, un moment raté et un moment qui posait question. Ces débriefings ont été enregistrés par Cathy Cohen et transcrits par Kim Szczot.

4.3.6 Interviews après le semestre

A la fin de toutes les séances, Julie Vidal a mené des interviews en français avec des étudiants dublinois à Dublin, et Marie Josée Hamel a recueilli des interviews menées auprès des apprentis-enseignants français. Sylvie Thouësny les a déposés dans un dossier partagé.

4.3.7 Choix et anonymisation des participants : enjeux éthiques

Toutes les données dont nous disposons sont sensibles. Le consentement des participants pour être enregistrés et faire partie d'une expérience est indispensable (Ten Have, 1999), et toutes les personnes doivent donner l'autorisation :

- d'être filmés ;
- que les enregistrements soient utilisés à des fins de recherche ;
- que les enregistrements soient publiés ou montrés au public.

Des formulaires de consentement ont été conçus par Nicolas Guichon et François Blin et distribués à tous les participants à Lyon et à Dublin. Pendant les interactions, au mois d'octobre 2013, les apprentis-enseignants ont rempli un formulaire leur rappelant que les interactions s'inscrivaient dans le cadre d'un projet de recherche et que les données recueillies étaient susceptibles d'être exploitées et partagées à des fins de recherche. Tous les apprentis-enseignants ont signé ce document, nous accordant ainsi leur permission de les inclure dans le corpus.

En décembre 2013, après les interactions, les apprenants dublinois ont répondu à un formulaire de consentement en anglais beaucoup plus élaboré que le formulaire de consentement qu'avaient signé les apprentis-enseignants français. Après avoir rappelé aux apprenants qu'ils avaient participé à six interactions susceptibles d'alimenter la recherche en didactique des langues, le questionnaire a sollicité leur accord sur les points suivants :

- les données anonymes (captures d'écran, enregistrements audiovisuels, citations) soient inclues dans le corpus, corpus susceptible d'être partagé entre chercheurs ;
- des extraits du corpus soient présentés à des conférences sans que le visage ne soit flouté ;
- des extraits du corpus soient présentés à des conférences à condition que le visage soit flouté ;
- des extraits du corpus soient publiés dans des articles scientifiques sans que le visage ne soit flouté ;
- des extraits du corpus soient publiés dans des articles scientifiques à condition que le visage soit flouté ;
- des citations anonymes soient présentées dans des publications scientifiques ;
- ils acceptent d'être interviewés dans le futur.

Contrairement au formulaire de consentement destiné aux apprentis-enseignants, qui était de type « tout ou rien », ce questionnaire destiné aux apprenants offrait à ces derniers la possibilité de nuancer leur consentement en répondant « oui » à certaines questions et « non » à d'autres. Ceci nous a permis dans un premier temps de savoir qui était éligible d'être inclus dans le corpus (tous ceux qui ont refusé que leurs données anonymes soient exploitées n'ont pas été inclus dans le corpus). Parmi les dix-neuf étudiants dublinois qui participaient aux interactions, cinq ont refusé de faire partie de l'étude ou n'ont pas répondu au questionnaire. Les autres ont donné leur accord à des degrés différents. Certains ont accepté que leurs visages soient présentés en clair lors des présentations, mais voulaient qu'ils soient floutés pour des publications. Ce besoin de floutage a amené à réfléchir à différentes façons d'anonymiser l'image de ces participants.

L'anonymisation des participants doit être systématique pour protéger l'identité des participants (Ten Have, 1999). Tous les participants ont donc reçu des pseudonymes, choisis par Ciara Wigham, quelles que soient leurs réponses aux formulaires de consentement. Les pseudonymes s'accordaient avec le genre des vrais prénoms et ont été choisis de manière à respecter la culture d'origine des participants. En plus d'accorder aux participants des pseudonymes, Ciara Wigham a créé un schéma pour renommer tous les fichiers en remplaçant les noms des participants par des chiffres. Ce document accompagnateur a été créé pour d'aider les membres du groupe à déchiffrer les noms des fichiers. Sylvie Thouësny a créé un script pour anonymiser automatiquement les fichiers texte. Ainsi, les messages envoyés par

clavardage ont été automatiquement et systématiquement anonymisés. L'anonymisation des vidéos sera discutée ultérieurement quand nous aurons expliqué la construction de celles-ci.

4.4 Construction des vidéos

Comme nous avons évoqué dans ce chapitre, un grand travail de design et de recueil avait déjà été effectué avant notre arrivé à Lyon. Cependant, les données audiovisuelles n'étaient pas exploitables car elles n'étaient pas dans des formats lisibles par des logiciels d'annotation. Notre contribution à la construction du corpus est donc celle de la reconstruction des vidéos que nous analysons. Si le recueil des plans de séances est aussi simple que le téléchargement de fichiers PDF, le recueil des vidéos s'est révélé compliqué et a nécessité la collaboration de plusieurs membres de notre laboratoire.

Lors des interactions, les flux audiovisuels et le clavardage pour chacun des participants ont été enregistrés. Quand nous sommes arrivé à Lyon en 2014, ces traces étaient accessibles à deux endroits. Premièrement, ces vidéos étaient visibles dans le salon de rétrospection de Visu susmentionné. Secondement, des enregistrements extraits par Sylvie Thouësny étaient disponibles dans un dossier partagé. Au début de la construction du corpus, ces deux lieux de stockage (Visu et Box[3]) nous offraient deux façons différentes de recueillir les données audiovisuelles. Nous allons les décrire en faisant le point sur les avantages et inconvénients de chaque méthode de recueil, ainsi que les décisions que nous avons prises en conséquence, et les changements de nos méthodes de recueil en fonction de l'évolution de nos compétences techniques.

4.4.1 Les traces disponibles depuis Visu

Dans le salon de rétrospection de Visu, les flux audiovisuels de chaque participant (une webcam et un micro par participant) sont visibles et audibles en cliquant sur « lecture ». Sur la Figure 7 du premier chapitre en page 31, l'image 1 est celle de l'enseignante, et les deux images à droite (2) sont celles des apprenants. Les images que l'on voit lorsqu'on se connecte au salon de rétrospection de Visu sont des reconstructions faites à partir de fichiers audiovisuels qui ont été crées automatiquement lors de l'interaction et stockés sur un serveur. Ces fichiers audiovisuels sont resynchronisés par Visu et s'affichent en grande taille ou en petite taille selon le choix de l'utilisateur (en l'occurrence, c'est l'image de l'enseignante qui s'affiche en grande taille). Ces données ne sont accessibles à l'utilisateur que de manière

[3] Box est une plateforme de stockage en ligne.

indirecte car il est impossible de les télécharger directement pour les visionner ou les traiter à l'aide d'autres logiciels.

Pour pallier l'impossibilité d'exporter les données de Visu, nous avons recouru à des captures dynamiques d'écran. Nous avons lancé un enregistreur d'écran (QuickTime)[4] et appuyé sur la touche lecture de Visu afin de capter toute l'interaction. QuickTime nous permet de sélectionner une partie de l'écran à enregistrer. Puis, nous avons enregistré un interactant à la fois afin d'avoir un fichier audiovisuel pour chacun. L'avantage de cette méthode est qu'*a priori*, les flux sont déjà synchronisés temporellement par Visu, ce qui veut dire qu'il n'y a pas *a priori* de resynchronisation à faire. Pour une interaction durant 40 minutes, trois captures dynamiques d'écran (une par participant) nous fournissent trois fichiers audiovisuels de 40 minutes qui sont synchronisés temporellement. Ce travail nécessiterait théoriquement 120 minutes de travail.

Cette façon de recueillir des données audiovisuelles a plusieurs inconvénients. La qualité audiovisuelle se dégrade à chaque fois que l'on fait une nouvelle capture d'écran (de même que la qualité se dégrade lorsqu'on fait une photocopie d'une photocopie) et les premiers et derniers moments de l'interaction sont parfois manquants. Les flux audiovisuels sont parfois désynchronisés sur Visu, principalement vers la fin des interactions (on entend la réponse à une question avant que celle-ci ne soit posée), et il est très difficile de corriger un décalage apparaissant au milieu de l'interaction si on ne dispose que d'un seul fichier par interactant. En outre, Visu affiche les noms des interactants sur les fenêtres, ce qui alourdit le travail d'anonymisation (cf. *infra* section 4.5 en page 128). Enfin, effectuer des captures d'écran en temps réel pour chaque participant est très chronophage : il faut 120 minutes pour capter une interaction de 40 minutes avec trois participants.

Le salon de rétrospection de Visu a été disponible en ligne pendant plus d'un an avant de disparaître définitivement pour des raisons de coûts et de dysfonctionnements institutionnels. Faute de compétences techniques, nous avons démarré le recueil du corpus en utilisant cette méthode qui avait été développée lors de notre mémoire de Master 2.

4.4.2 Les traces disponibles depuis Box

Lors des interactions, un fichier audiovisuel en format FLV[5] a été créé lors de chaque connexion de chaque interactant. En effet, un interactant qui se connecte et se déconnecte

[4] Ce logiciel permet d'enregistrer tout ce qui se passe sur l'écran ou sur une partie de l'écran de l'ordinateur.
[5] FLV est un format vidéo conçu pour Adobe Flash Player.

trois fois lors d'une interaction, ce qui arrive très souvent pour diverses raisons techniques, va avoir trois fichiers distincts. Ces fichiers ont été extraits par Sylvie Thouësny et stockés dans un dossier partagé sur Box auquel tous les membres du projet ISMAEL avaient accès. Contrairement au salon de rétrospection de Visu, ces fichiers FLV peuvent être téléchargés directement sur un ordinateur. Aucune capture d'écran n'est requise.

Cette méthode de recueil a plusieurs avantages. Nul besoin de capture d'écran à refaire : il suffit de télécharger directement les fichiers. Chaque participant a sa propre image et son son audio. La qualité audiovisuelle n'est pas dégradée car il s'agit de fichiers bruts d'origine. Enfin, les noms des participants ne s'affichent pas dans les fenêtres, ce qui facilite l'anonymisation du corpus.

L'inconvénient de cette méthode et qu'il faut reconstruire, à partir d'extraits, des vidéos complètes qui ressemblent à celles que l'on voit dans le salon de rétrospection de Visu. Ceci nécessite des compétences de synchronisation et de montage car le nombre de fichiers par séquence va de deux (pour une interaction avec deux interactants, un fichier chacun) à dix-huit (pour une interaction avec trois interactants ayant respectivement huit, cinq et cinq fichiers chacun). Il ne s'agit pas simplement de les coller l'un après l'autre car ces fichiers ne s'alignent pas temporellement les uns avec les autres. Ceci est dû au fait que tous les participants ne se connectent et ne se déconnectent pas en même temps et pour les mêmes durées. Par conséquent, il y a des vides (durée de temps où l'interactant n'est pas connecté et on ne voit rien sur l'écran) avant et après certains extraits.

4.4.3 Choix entre les deux façons de recueillir le corpus

Idéalement, afin d'avoir un corpus de la meilleure qualité possible, il faudrait construire tout le corpus à partir des fichiers FLV dans la Box en les synchronisant sur des logiciels de montage audiovisuel. Or, au début du recueil du corpus, nous ne disposions pas des compétences techniques nécessaires pour reconstruire des vidéos à partir de ces fichiers bruts. Il a fallu donc repérer les caractéristiques techniques de chaque séance afin de décider entre les deux options évoquées ci-dessus. Si une interaction avait peu de fichiers bruts et que ceux-ci étaient déjà synchronisés, nous les montions en utilisant QuickTime Pro. Ce logiciel nous permettait d'attacher ensemble des fichiers déjà synchronisés mais ne nous permettait pas de faire la synchronisation. Si par contre le nombre de fichiers bruts était trop important et/ou pas synchronisés et que le salon de rétrospection de Visu représentait fidèlement

l'interaction, nous effectuions des captures dynamiques d'écran. Nous avons donc construit un tableau pour chaque séance afin de repérer, à partir des extraits, les caractéristiques techniques sur Visu et sur Box.

Semaine n°
Nom du tuteur
Nb. d'interactants

Sur Visu
Durée totale :
Synchronisation au début :
Synchronisation 10m :
Synchronisation 20m :
Synchronisation 30m :
Synchronisation 40m :
Synchronisation fin :

Sur Box
Nb. de fichiers tuteur :
Durée totale tuteur : =
Nb. de fichiers A1 :
Durée totale A1: =
Nb. de fichiers A2 :
Durée totale A2 : =
Première image correspond : tuteur : A1 : A2 :
Durées correspondent (T=A1=A2=Visu) :

Figure 2 : Caractéristiques des interactions

Cette fiche nous a permis de faire état de la qualité de synchronisation sur Visu et du nombre de fichiers bruts à synchroniser le cas échéant. Après investigation, nous avons procédé à la construction d'à peu près la moitié du corpus par les fichiers FLV, les autres interactions se révélant trop compliquées à synchroniser à partir de ceux-ci.

Pour synchroniser les flux FLV (option 2), nous avons sollicité l'aide de la Cellule Corpus Complexes (CCC). La CCC se trouve au sein su laboratoire ICAR et « vise à la mutualisation des pratiques sur les corpus au sein du laboratoire pour traiter des phénomènes "linguistiques" dans toute leur complexité multimodale et pluridisciplinaire » (CCC). Justine Lascar, une ingénieure de la CCC, nous a créé certaines séances au début sur Final Cut Pro, et nous a dispensé une formation initiale sur ce logiciel. C'est grâce à cette formation que nous avons pu devenir autonome. Après avoir acquis cette compétence technique, nous avons décidé de reconstruire la majorité du corpus en utilisant ce logiciel.

4.4.4 Synchronisation des Flux sur Final Cut Pro

Nous allons décrire la façon dont nous avons synchronisé les séances sur Final Cut Pro, logiciel de traitement audiovisuel qui est utilisé par des professionnels pour créer des films et dont la prise en main est compliquée. Ce logiciel possède de nombreuses fonctionnalités diverses, mais pour le projet ISMAEL nous utilisons principalement la fonctionnalité de la synchronisation de multiples flux audiovisuels.

Avant de traiter les fichiers sur Final Cut Pro, il a fallu les convertir en format MP4[6] car le format FLV n'est pas lisible par Final Cut Pro. Nous avons utilisé HandBrake, logiciel très adapté à cette tâche.

Nous avons synchronisé deux flux à la fois sur Final Cut Pro. Nous avons pris l'exemple suivant : nous avons voulu construire l'interaction de Samia (apprentis-enseignante), Angela et Sean (apprenants) pour la semaine 3. Il y avait 13 fichiers FLV (2 pour Angela, 6 pour Samia et 5 pour Sean) qui ont été convertis en fichiers MP4 par HandBrake.

Il a fallu sélectionner samia_s3_1 et sean_s3_1, parce que les durées étaient similaires et parce qu'Angela était absente au début de l'interaction. Puis, il a fallu ouvrir ces deux flux dans le « plan synchronisation ». Final Cut Pro tente d'abord de synchroniser les flux automatiquement, mais cela ne fonctionne pas systématiquement. Pour la synchronisation automatique, Final Cut Pro essaie de synchroniser les flux automatiquement en se basant sur l'audio. Il essaie de repérer un bruit qui se trouve dans les deux flux au même moment afin de les aligner. Malheureusement, le son de l'interlocuteur est souvent très bas, donc difficile à entendre. Ensuite, comme les autres interactants ne sont pas dans la même salle et n'ont pas les mêmes bruits de fond, le son se décale à cause du dispositif, ce qui fait que même si le son de l'interlocuteur est audible, la synchronisation exacte est impossible. Nous avons donc synchronisé les flux manuellement en écoutant nous-même et en posant des marqueurs ainsi :

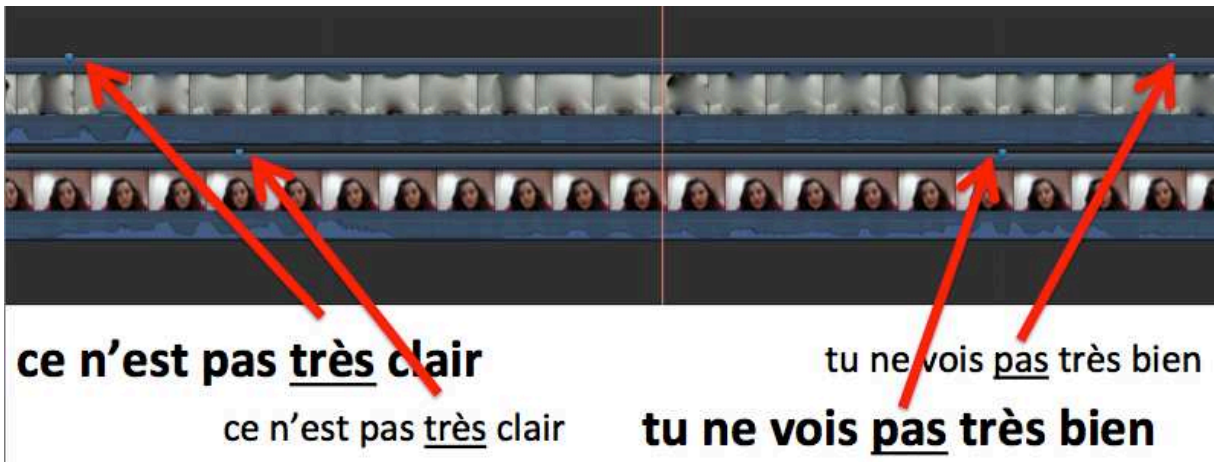


Figure 3 : Alignement des flux audiovisuels sur Final Cut Pro

La bande en haut (voir Figure 3) représente le flux de Sean, et celle d'en bas représente celui de Samia. Le temps passe de gauche à droite. Final Cut Pro nous permet de déplacer le flux de Sean de la gauche vers la droite par rapport à celui de Samia. Ainsi, nous

[6] MP4 est un format vidéo fait pour être lu en streaming.

pouvons synchroniser temporellement ces deux flux audiovisuels. Pour ce faire, nous repérons un mot qui est audible dans les deux flux. C'est Sean qui dit « ce n'est pas très clair », et dans son flux audio il est très facile de l'entendre. Dans le flux de Samia, le son est plus bas car il sort de son casque et est recapté par son micro. On pose un marqueur au moment où on entend le mot « très ». Il serait tentant de les aligner sur ce mot, mais nous ne pouvons pas le faire parce qu'il y a un décalage dans les deux directions. Pour prendre en compte ce décalage, on fait la même chose mais dans l'autre direction. Cette fois-ci, c'est Samia qui dit « tu ne vois pas très bien », phrase qu'on entend également comme bruit de fond dans le flux de Sean. On pose un marqueur là où on entend le mot « pas ». Après avoir posé les marqueurs, on égalise la distance entre ces deux points en déplaçant la piste d'au-dessus soit à gauche soit à droite. Ceci nous donne une représentation la plus neutre possible de l'interaction. Ce décalage a des implications méthodologiques qui seront éclaircies à la fin de ce chapitre.

Après chaque épisode de synchronisation, nous ajoutons les flux synchronisés au projet global (au sens Final Cut Pro), qui ressemble à ceci :

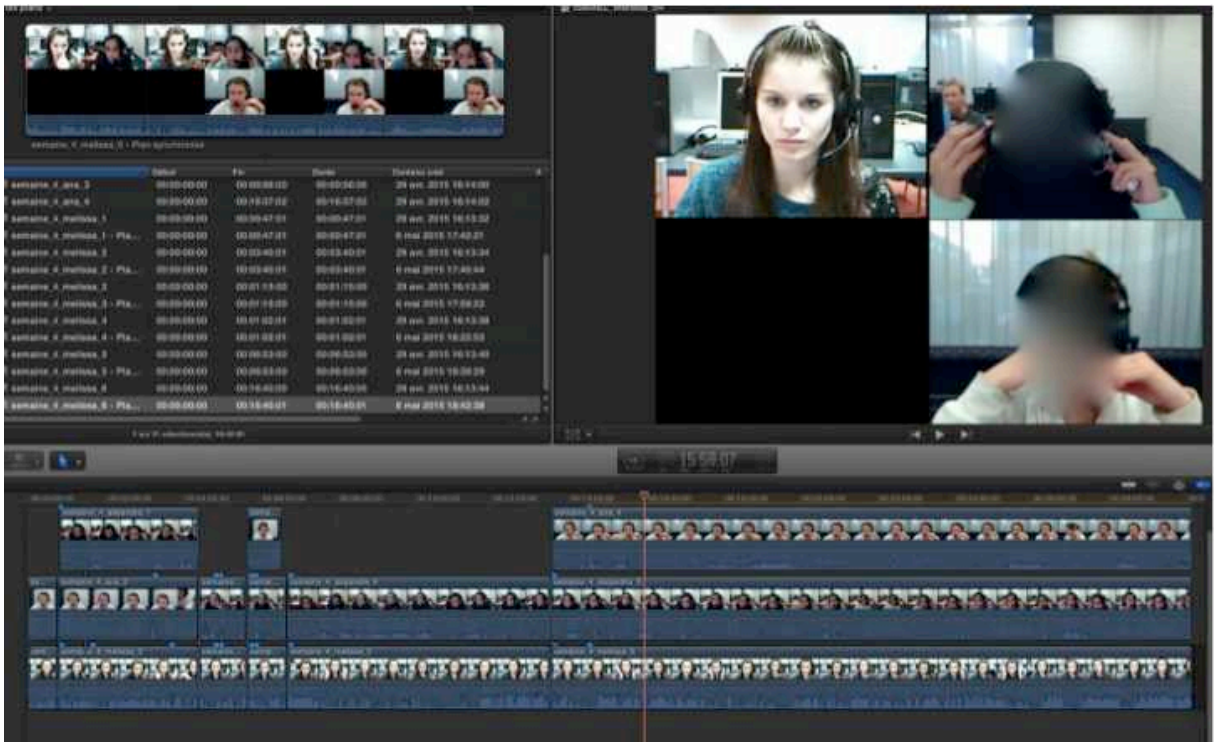


Figure 4 : Projet global sur Final Cut Pro

Dans cette capture d'écran de Final Cut Pro (Figure 4), nous voyons six sections de gauche à droite et 15 fichiers. Pour rappel, ces 15 fichiers étaient d'abord des fichiers FLV, puis des fichiers MP4. Après la synchronisation, tous les fichiers sont alignés. La dernière étape

consiste à exporter un fichier audiovisuel par interactant en format MP4, qui a exactement la même durée. Final Cut Pro remplit les espaces vides entre les fichiers en insérant automatiquement un espace noir en fonction du temps de déconnexion. Après avoir créé un fichier audiovisuel par interactant pour chaque séance, nous les avons exportés en utilisant Compressor, un logiciel pour Mac. De cette façon, nous avons créé un fichier audiovisuel par interactant par séance.

4.4.5 Montage final sur QuickTime Pro

Après avoir crée les fichiers audiovisuels individuels (un par interactant), que ceux-ci aient été créés en faisant des captures dynamiques de l'écran de Visu ou en synchronisant les fichiers bruts sur Final Cut Pro, nous avons dû combiner les flux avant de les transcrire afin de voir tous les interactants ensemble comme on les verrait dans le salon de rétrospection de Visu. Pour faire cette manipulation, nous utilisons QuickTime Pro, un logiciel qui permet de faire un multiscope[7]. Dans la figure ci-dessous, on voit les trois flux avant et après le montage sur QuickTime Pro. Les trois flux en haut ont été reconstitués sur Final Cut Pro. Ils ont été combinés par la suite sur QuickTime Pro pour former une seule vidéo, le multiscope, en bas.

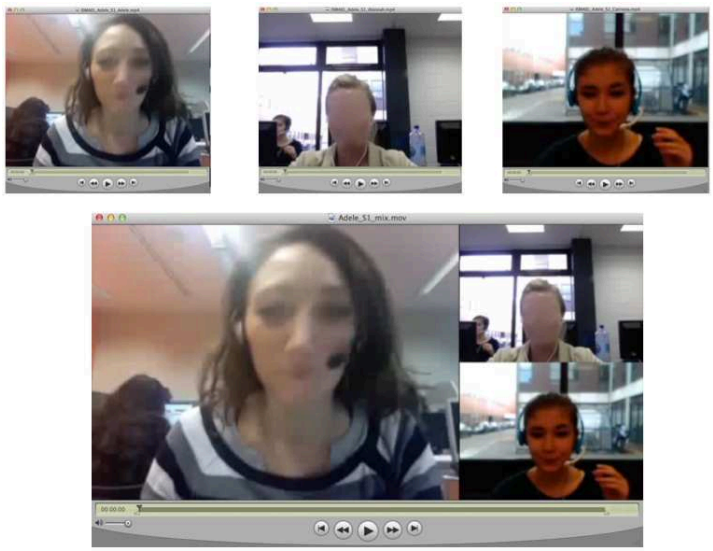


Figure 5 : Montage final sur QuickTime Pro

A l'instar de l'affichage dans le salon de rétrospection de Visu et parce que notre thèse se concentre sur les compétences des enseignants, nous avons mis l'image de l'apprenti-enseignant deux fois plus grande que celles des apprenants. Après avoir effectué le montage, nous avons utilisé QuickTime Pro pour exporter un fichier vidéo en format MP4 qui contient toutes les fenêtres vidéo, et un fichier son en format .wav pour chaque participant. C'est pour

[7] Le multiscope nous permet de voir deux ou trois interactants à la fois sur la même vidéo.

cette raison qu'il nous a fallu construire un fichier par interactant. Dans l'exemple ci-dessus (Figure 5), 4 fichiers ont été créés : un fichier vidéo et trois fichiers son. Ces fichiers vont nous permettre de transcrire de manière fine toute l'interaction. En effet, les fichiers .wav nous permettent de visualiser le son sur le logiciel ELAN, ce qui rend plus facile la segmentation et la transcription de la parole orale, et le fait d'avoir de multiples fichiers son nous permet de sélectionner l'interactant que l'on veut écouter à un moment donné.

Pour résumer, les données se trouvaient à deux endroits différents : sur Visu et sur Box. Ces deux endroits nous ont offert deux possibilités de recueillir le corpus. Au début de la construction du corpus, nous avons favorisé le recueil par Visu grâce à sa simplicité. Plus tard, après avoir approfondi nos compétences techniques sur l'usage de Final Cut Pro, nous avons reconstruit la plupart du corpus par la seconde méthode, ce qui nous a fourni un corpus de meilleure qualité. Les fichiers audiovisuels crées par Final Cut Pro ont été ensuite combinés sur QuickTime Pro afin de voir trois interactants à la fois, à l'instar du salon de rétrospection de Visu. C'est ainsi que nous avons reconstitué les séances de notre corpus qui avaient été fabriquées auparavant par d'autres moyens moins précis. Notre corpus est désormais plus clair et synchronisé. Le tableau ci-dessous schématise les différents logiciels dont nous nous sommes servi afin de construire la partie audiovisuelle du corpus.

Logiciel	**Manipulation technique**	**Compétences requises**	**Difficulté**
QuickTime	Effectuer des captures d'écran de Visu	Sélectionner la partie de l'écran à enregistrer	Facile
HandBrake	Convertir des fichiers FLV en format MP4	Sélectionner les fichiers	Facile
Final Cut Pro (FCP)	Synchroniser les fichiers audiovisuels	Manipuler les outils de synchronisation dans la timeline de Final Cut Pro, construire une vidéo complète à partir de morceaux, créer une vidéo par participant	Difficile au début
Kompressor	Exporter les vidéos synchronisées sur FCP, une vidéo par participant	Régler les paramètres d'exportation	Facile
QuickTime Pro	Combiner deux ou trois vidéos déjà synchronisées pour faire un multiscope, exporter un fichier vidéo par séance et un fichier audio par participant	Régler la taille des images manuellement en comptant le nombre de pixels, régler les paramètres d'exportation	Moyen

Figure 6 : fonctions de chaque logiciel et les compétences requises

4.4.6 Un point de vue « observateur »

Comme nous l'avons mentionné, notre façon de synchroniser les flux audiovisuels et de les combiner a des implications méthodologiques, à la fois pour le son et pour la vidéo.

Nous avons effectivement créé un point de vue observateur, et aucun des interactants ne voit l'interaction de ce point de vue. Pour mieux illustrer ce phénomène, imaginons une situation dans laquelle il y ait un apprenant et un enseignant qui interagissent en visioconférence, et qu'il existe un décalage non variable d'une seconde dans chaque direction (enseignant vers apprenant et inversement)[8]. Si l'enseignant énonce le mot « je », l'apprenant va l'entendre une seconde plus tard, et inversement. Si les deux participants énoncent en même temps le mot « je », ce ne sera perçu comme un chevauchement par aucun des interactants car le « je » de l'autre sera entendu une seconde plus tard. Si l'enseignant dit « je » et que l'apprenant le dit une seconde plus tard, ceci sera perçu comme un chevauchement par l'apprenant mais pas par l'enseignant, ce dernier entendant le « je » deux secondes après l'émission du sien. Il est donc impossible pour l'observateur dans ce cas de se positionner à la fois du point de vue de l'enseignant et de celui de l'apprenant. Nous avons donc décidé de faire un compromis en nous plaçant entre les deux, en égalisant les décalages et en créant ainsi un point de vue observateur. Grâce à ce nouveau point de vue, si les participants énoncent « je » en même temps, le chercheur va percevoir un chevauchement qui ne sera perçu comme tel par aucun des interactants. Ainsi, les chevauchements que l'on transcrit sont à prendre en compte avec prudence.

En ce qui concerne le visuel, nous avons fait le choix de fixer l'image de l'enseignant plus grande que celle des interactants afin que nos vidéos ressemblent à ce que l'on voyait dans le salon de rétrospection de Visu et cela facilite l'annotation sur ELAN (cf. *infra* section 4.6.1 en page 130). Cette façon d'organiser les fenêtres nous permet de visualiser trois interactants en même temps et de nous concentrer surtout sur le corps et le visage de l'enseignant.

4.5 Anonymisation des images

Nous avons évoqué plus haut l'importance de l'anonymisation des données, et la procédure par laquelle des pseudonymes ont été accordés aux participants. Il est question ici d'anonymiser les images de ceux qui ont répondu négativement à certaines questions sur le

[8] Dans notre corpus, il y a souvent trois interactants au lieu de deux, et les décalages sont variables, allant de moins d'une seconde à plusieurs secondes.

formulaire de consentement. Nous allons décrire les enjeux éthiques que nous avons considérés avant, pendant et après la construction de la partie audiovisuelle du corpus.

Au début de la construction de la partie audiovisuelle du corpus, quand certaines vidéos étaient toujours issues des captures dynamiques d'écran de Visu, nous avons posé des bandeaux noirs sur les noms des participants qui se trouvaient en haut à droite de toutes les vidéos (Figure 7). Plus tard, quand nous avons reconstitué notre corpus en utilisant uniquement les fichiers « bruts » en format FLV se trouvant dans Box, ces bandeaux noirs sont devenus obsolètes car les noms des participants ne s'affichaient plus.

Sur les douze apprenants que nous avons retenus pour le corpus, huit ont souhaité que leurs visages soient floutés dans des publications. Pour flouter une image statique, nous avons utilisé la fonction « retoucher » de l'application iPhoto[9]. Il est plus compliqué de flouter une vidéo. La manière la plus simple est de poser un carré noir sur le visage de l'apprenant de la même manière que nous avons posé des rectangles noirs sur les prénoms des participants dans les vidéos issues de Visu. L'inconvénient est qu'on ne voit plus le visage. Une meilleure façon d'anonymiser le visage est de poser une image statique flouté du visage sur l'image du participant dans la vidéo. L'inconvénient de cette technique est que l'image floutée de l'interactant reste figée et masque en permanence l'image claire et dynamique d'origine. Il est possible de faire du floutage dynamique sur Final Cut Pro, de façon à ce que l'image bouge et reste floutée en même temps, mais nous avons décidé de ne pas procéder à cette technique car elle est chronophage et parce qu'une image floutée est incapable de montrer les expressions faciales. L'image ci-dessous illustre ces deux phénomènes que nous venons d'expliquer : les rectangles noirs sur les noms des participants et le floutage statique de l'image d'Aiden.

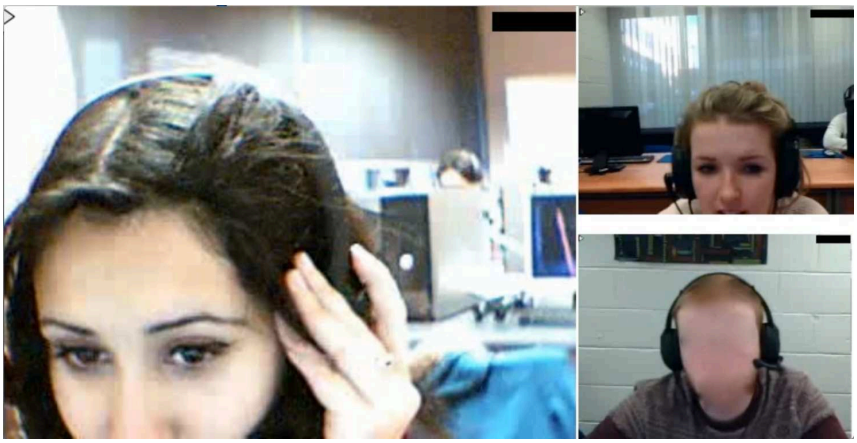


Figure 7 : Bandeaux noirs et visage flouté

[9] iPhoto est un éditeur de photos.

Une solution qui a été envisagée[10] ultérieurement est la « cartoonisation » des images des acteurs (voir les images des apprenants sur la Figure 8). L'effet de « cartoon » que peut appliquer sur une vidéo le logiciel VLC a deux grands avantages par rapport au floutage. Principalement, il rend anonyme les acteurs sans pour autant en masquer les expressions faciales, gestes et mouvements de corps. Ensuite il est très facile d'appliquer cet effet sur une vidéo en utilisant VLC sans avoir des compétences en traitement vidéo. Quant aux inconvénients, il faut recréer le morceau vidéo auquel on souhaite appliquer cet effet, car il est impossible de ne cartooniser qu'une partie de l'écran. On doit cartooniser toute la vidéo puis couper les images des participants qu'on souhaite anonymiser, les superposer sur la vidéo d'origine, et exporter une toute nouvelle vidéo en utilisant QuickTime Pro. Pour ces raisons, nous n'appliquons cette mesure de floutage qu'aux morceaux présélectionnés destinés à des communications scientifiques.

4.6 Transcription des données

Un corpus est plus rentable scientifiquement s'il est facile d'accès et transcrit. Nous allons expliquer dans cette section la façon dont nous avons transcrit et mis à la disposition de la communauté de chercheurs le corpus ISMAEL.

4.6.1 Transcription du corpus

Après avoir créé les fichiers audiovisuels du corpus, nous les avons importés dans le logiciel ELAN (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006) pour les transcrire. Voici ce à quoi ressemble ELAN :

[10] C'est Justine Lascar de la CCC qui nous a proposé cette solution.

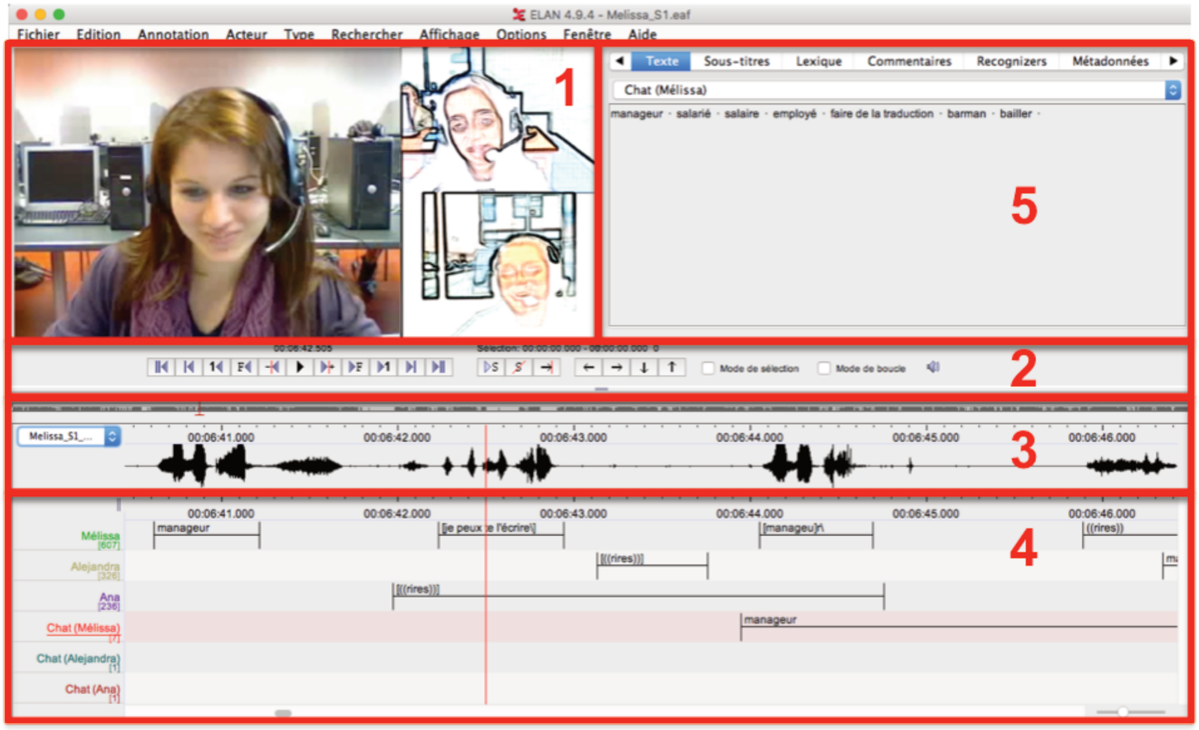


Figure 8 : ELAN

Les vidéos que nous avons créées sont importées dans le logiciel ELAN et apparaissent en haut à gauche (1). Les boutons de lecture se trouvent en dessous (2). Les fichiers son en format .wav nous permettent de visualiser le son de chaque interactant et de sélectionner le moment exact où chacun parle afin de transcrire la parole (3). Les pistes d'annotation (4) nous permettent d'annoter la parole orale, l'usage du clavardage, les phénomènes multimodaux (cf. chapitre 3), et de délimiter les séquences d'explication lexicale.

Pour chaque phénomène annoté, nous avons créé une piste d'annotation. Les pistes d'annotation pour la parole ont été transcrites de manière collective (voir annexe 4 en page 387), tandis que les pistes d'annotation pour les phénomènes multimodaux ont été créées par nous-même pour répondre à nos propres questions de recherche. L'annotation sera détaillée dans le prochain chapitre.

En 2014, 2015 et 2016 nous avons ajouté de nouvelles séances au corpus au fur et à mesure en créant les fichiers audiovisuels nécessaires à la transcription et en les annotant sur ELAN. Après les avoir créés, nous avons réparti le travail de transcription de la parole orale. Trois étudiants en master[11] ont été embauchés pour nous aider à transcrire le corpus. Nous les avons formés à utiliser le logiciel ELAN et nous les avons familiarisés avec la convention

[11] Thomas Debay, Aline Pellicier, Angèle Guyard

ICOR[12] (Groupe ICOR, 2013). Aucune transcription n'est neutre (Mondada, 2000), mais nous avons sélectionné cette convention de transcription qui est fréquemment utilisée au sein du laboratoire ICAR afin que nos transcriptions puissent être réemployées par un grand nombre de chercheurs. Au fur et à mesure de la création des fichiers audiovisuels nous avons distribué ceux-ci aux transcripteurs afin qu'ils les transcrivent sur ELAN.

Les messages envoyés par clavardage étaient stockés dans un document en format .xml contenant des milliers de lignes de code. Nous avons collaboré avec un développeur[13] qui a créé un script qui a extrait automatiquement tous les messages envoyés, ainsi que la date et l'heure correspondantes. Les messages ont été intégrés aux fichiers de transcription sur ELAN.

Ultérieurement, d'autres étudiants en master ont transcrit chacun une ou deux séances[14]. Julie Vidal, dont la thèse porte également sur ce corpus, a transcrit quelques séances. Après ce travail de construction et de transcription, notre corpus est constitué de 28 séances pour un total d'environ 15 heures et 23 minutes. La figure ci-dessous, que nous avons déjà présentée dans le premier chapitre, donne une vue synoptique du corpus ISMAEL dans son état présent[15] :

[12] Cette convention a été choisie pour sa polyvalence.

[13] Matthieu Décorde

[14] Cela leur a donné la possibilité d'accès au corpus pour analyser certaines parties pour leurs mémoires de Master 2.

[15] Aparajita Dey, dont le titre provisoire de la thèse est « Designed and Emerging Affordances in Tutor-Learner Videoconferencing & Multimodal Interactions for Second Language Development: An Ecological Approach » continue à élargir le corpus en montant et en transcrivant de nouvelles séances.

Apprentis-enseignants	Apprenants		Séance 1 15 octobre Les droits du travail	Séance 2 22 octobre Expériences professionnelles	Séance 3 12 novembre Se préparer pour un stage à Reims	Séance 4 19 novembre Gestion d'un projet	Séance 5 26 novembre Mettre en place un projet	Séance 6 3 décembre Entretien d'embauche
Adèle* 	Alannah 	Catriona* 	☑ 43m48s	☑ 40m55s	☑ 33m03s	☑ 34m27s	☑ 41m04s	☑ 35m55s
Emilie* 	Aiden 	Fiona 	☑ 22m50s	☑ 33m00s		☑ 25m37s		☑ 24m12s
Samia* 	Sean* 	Angela* 	☑ 30m20s	Sean absent ☑ 34m34s	⊙ ☑ 31m50s	☑ 35m20s	☑ 39m46s	x
Etienne* 	Conor 	Sophie 				☑ 20m57s		
Mélissa* 	Ana* 	Alejandra 	☑ 42m11s	☑ 39m24s	☑ 37m03s	☑ 30m13s	☑ 32m28s	☑ 35m53s
Séverine* 	Naomi 			x		☑ 34m47s		
Victor* 	Liam 		☑ 33m25s	Séance avec Conor ☑ 27m46s	☑ 29m18s	☑ 34m00s	⊙ ☑ 29m01s	x

☑ Séance transcrite
x Séance non disponible

* Entretien
⊙ Vue du hors champ disponible
Temps total transcrit : 15h, 23m, 07s

Planning de la séance
Débriefing le lendemain

Figure 9 : Vue synoptique du corpus ISMAEL

Dans les deux premières colonnes à gauche se trouvent les sept apprentis-enseignants et leurs apprenants respectifs, ainsi que les pseudonymes qui leur ont été donnés. Les apprenants souhaitant rester anonymes ont été cartoonisés afin que les chercheurs travaillant sur ce corpus puissent se référer rapidement aux règles éthiques. Les six colonnes suivantes représentent les six séances qui ont eu lieu hebdomadairement. La date et le thème de chaque séance sont indiqués en haut. Les cases en dessous contiennent les durées de chaque séance et sont cochées ou non selon que la vidéo de la séance a été recréée à l'aide de Final Cut Pro et transcrite ou non. Les cases vides ne font pas partie du corpus, et les cases avec une croix représentent des séances dont on aurait voulu qu'elles fassent partie du corpus mais dont les enregistrements n'ont pu être localisés. Un tableau récapitulant la durée, le nombre d'interactants, et le transcripteur de chacune des séances se trouve en annexe 4 en page 387 et permet de mettre en valeur le travail de construction et de transcription qui a été mené lors de la construction du corpus. Le temps requis pour créer une vidéo varie de 3 à 10 heures, et le temps requis pour transcrire la parole orale d'une séance est de 20 heures environ.

La figure ci-dessous récapitule les étapes entreprises afin de construire les séances qui font partie du corpus :

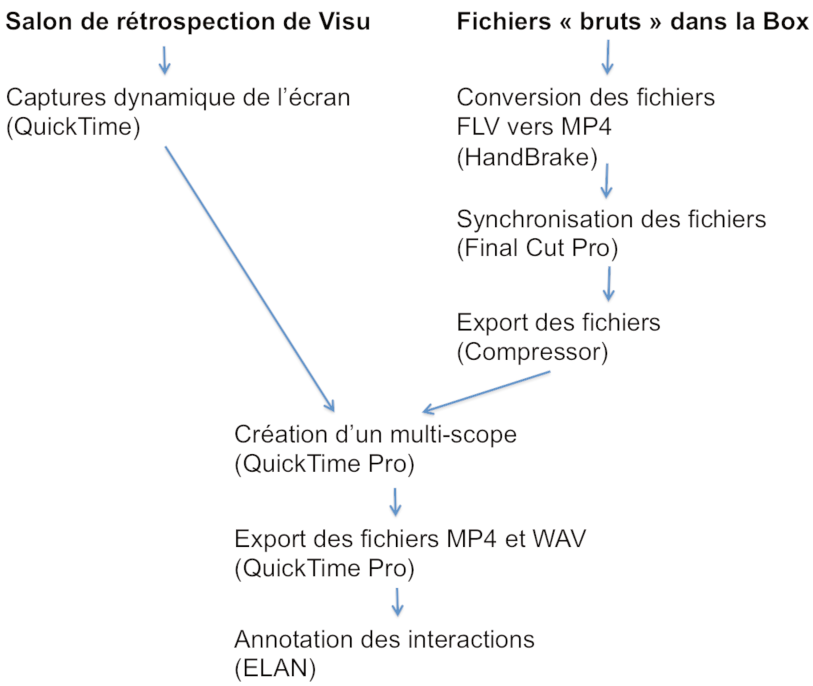


Figure 10 : Récapitulatif des opérations menées lors de la construction du corpus

4.6.2 Sélection des séances

Puisque la transcription d'une séance nécessite une vingtaine d'heures de travail, nous n'avons pas sélectionné les séances à ajouter au corpus de manière aléatoire. Les décisions n'ont pas été prises uniquement par nous-même, mais en collaboration avec les concepteurs du projet ISMAEL.

Le lecteur constatera que la totalité des séances de deux apprentis-enseignants ont été transcrites, et que toutes les séances de la quatrième semaine des cours ont été transcrites. Ces deux phénomènes relèvent de deux choix méthodologiques que nous avons faits. Nous avons voulu suivre quelques apprentis-enseignants tout au long du semestre (lignes horizontales de la Figure 9) afin de connaître leurs comportements et habitudes didactiques, repérer d'éventuels changements et évolutions, et faire émerger des typifications (habitudes propres à chacun qui se manifestent régulièrement). Cicurel (2011 : 129) explique que les typifications « sont le fondement de l'agir professoral ». En analysant les pratiques de ces enseignants, nous serons à même de voir émerger une grammaire constituée de pratiques typiques, une « grammaire des actions » (Cicurel, 2011 : 130). Nous avons donc choisi Adèle, Mélissa et Victor au début car nous avions déjà étudié leurs premières séances dans le cadre de notre mémoire de Master 2, et que ceux-ci nous paraissaient particulièrement intéressants. Adèle et

Mélissa pour leur expressivité, et Victor pour sa personnalité et culture différente des autres[16] et son corps et visage plutôt figés. Nous aurions aimé inclure les séances 6 de Victor et de Samia, mais les fichiers nécessaires à la construction de celles-ci sont introuvables.

Notre seconde décision a été de transcrire toutes les interactions de la séance 4 (ligne verticale de la Figure 9) afin de mener des études transversales. Pendant cette séance, tous les apprentis-enseignants font les mêmes activités, ce qui nous donne une occasion rare et précieuse de comparer l'agir professoral (Cicurel, 2011) de sept professeurs mis dans des situations similaires. Nous allons mener une comparaison interindividuelle dans le chapitre 9. Cette flexibilité du corpus est d'autant plus intéressante qu'il a été partagé avec d'autres chercheurs de différentes universités françaises dans le cadre du projet ISMAEL (Guichon & Tellier, 2017).

4.7 Partage du corpus

Quand le corpus a été rendu exploitable par sa construction et sa transcription, il a été partagé. Le partage du corpus a commencé bien avant que la construction de celui-ci ne fût terminée. Dans un dossier partagé dans Box qui avait été créé auparavant par Françoise Blin, Ciara Wigham a récolté divers documents (planning des cours, guide d'utilisation de Visu, enregistrements, bilans, entretiens, etc.) et les a déposé dans ce dossier. L'intérêt de partager un corpus, explique Guichon (2017b), est que celui-ci permet l'émergence de différentes perspectives et différents regards critiques sur les mêmes données ou sur les différentes sous-parties d'un corpus.

Afin de partager le corpus avec les chercheurs du projet ISMAEL, nous avons utilisé la plateforme Box, la même plateforme que nous avions utilisée auparavant pour télécharger les données « brutes » à partir desquelles la partie audiovisuelle du corpus a été construite. Ciara Wigham et Sylvie Thouësny ont créé des noms d'utilisateurs pour chacun des membres du groupe, et elles ont organisé les dossiers selon les niveaux de confidentialité. Cette hiérarchisation a permis d'avoir un espace privé pour les membres du projet et un espace public pour les autres chercheurs travaillant sur le corpus. L'espace public est divisé en niveaux aussi : données brutes, données annotées, données analysées, et extraits partagés. Nous avons organisé nous-même ce deuxième niveau, données annotées, en dossiers, un dossier pour chaque séance. Dans chaque dossier nous avons mis quatre ou cinq fichiers selon le nombre interactants. S'il y a trois interactants, il y a un fichier vidéo en format MP4

[16] Victor est d'origine québécoise.

(vidéo), trois fichiers WAV (un fichier son par interactant), et un fichier d'annotation en format EAF (ELAN). La navigation entre ces dossiers est susceptible de se révéler compliquée pour les utilisateurs, ce qui constitue la raison principale pour laquelle un portail a été créé par Thomas Debay et mis en ligne par Daniel Valéro pour en faciliter l'accès. En voici une capture d'écran :



Figure 11 : Portail du corpus ISMAEL

Sur ce site les utilisateurs peuvent cliquer sur la séance qu'ils souhaitent télécharger. Ainsi ils n'ont pas besoin de parcourir la liste des dossiers dans la Box. C'est en donnant le lien et le mot de passe du site aux membres du projet ISMAEL et aux étudiants en Master 2 que nous avons partagé ce corpus.

4.8 Type de corpus

De quel type de corpus s'agit-il ? Comme le rappelle Ten Have (1999), la frontière entre expérimental et écologique n'est pas opaque. C'est la raison pour laquelle il convient de rappeler le continuum méthodologique entre données contrôlées et données écologiques,

distinction que certains chercheurs se sont employés à nuancer (Tellier, 2013) :

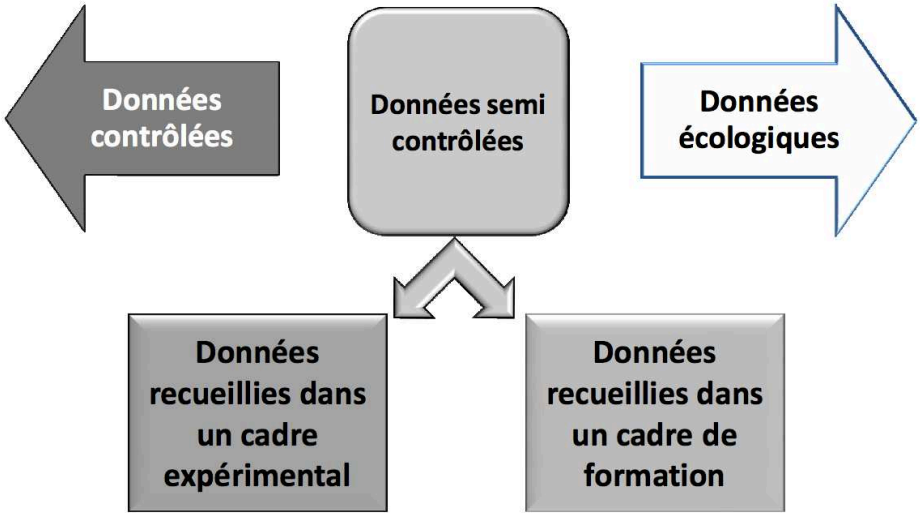


Figure : 12 : Continuum méthodologique (Tellier, 2013)

Comme l'explique Tellier (2013), d'un côté nous avons des données écologiques qui sont créées lorsqu'un chercheur place un enregistreur dans une salle de classe ou autre milieu dit naturel. Ces données riches et complexes sont utilisées pour des études qualitatives et des analyses descriptives qui ne font pas de généralisations et sont souvent des études de cas. D'autre part, des études expérimentales sont menées avec des données contrôlées et donnent des résultats quantitatifs. Pour recueillir ce deuxième type de données, certaines variables comme les conditions d'expérimentation et le protocole expérimental doivent être bien contrôlées et maîtrisées. Les données semi-contrôlées se trouvent alors à mi-chemin entre les données contrôlées et les données écologiques. Tellier (2013 : 5)[17] explique que

> « Le principe du corpus semi-contrôlé est que l'on recueille les productions (orales ou écrites) de différentes personnes à qui l'on donne les mêmes consignes. On peut donc contrôler certaines variables, telles que les personnes participant à l'étude et les conditions de passation. Les productions recueillies sont ainsi comparables entre elles. Les données récoltées peuvent mener à des analyses à la fois qualitatives et quantitatives. »

Ces corpus semi-contrôlés peuvent être recueillis soit dans le cadre d'une formation, soit dans un cadre expérimental. Tellier (2013 : 5-6) explique que « Dans le cadre de la formation, le chercheur n'intervient pas sur les consignes. Il recueille des productions que l'enseignant ou l'institution a sollicité » tandis que « dans le corpus semi-contrôlé recueilli dans le cadre d'une expérimentation, le chercheur détermine les conditions de production, les consignes, le lieu et le moment de passation ».

Dans notre corpus, les chercheurs ne sont pas intervenus avec des consignes précises. Avant chaque séance, les apprentis-enseignants se sont mis d'accord sur les activités à

[17] La pagination correspond au manuscrit auteur.

réaliser, les documents à partager, et les questions à poser. La durée des interactions avait été fixée à 45 minutes, mais en réalité les durées étaient variables. La tâche est contrôlée, mais les données ne le sont pas. Chacun fait une réinterprétation de la tâche et conduit l'interaction de sa propre façon. Le corpus devient moins contrôlé à cause des incidents techniques. Les nombreuses variables aléatoires que nous avons rencontrées nous ont obligé de préciser que le point où se trouve notre corpus sur le continuum méthodologique s'oriente plutôt vers un corpus écologique (Figure 13).

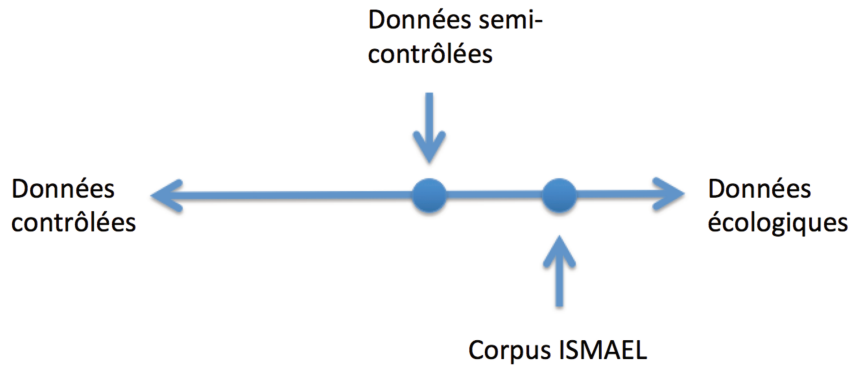


Figure 13 : Le corpus ISMAEL sur le continuum méthodologique

Notre corpus est plutôt écologique car chaque apprenant est différent et ne va pas répondre aux questions de la même façon. Certains apprentis-enseignants se trouvent en face de deux apprenants, tandis que d'autres n'en ont qu'un seul, ce qui aura un impact sur le déroulement de l'interaction, principalement au niveau de la gestion des tours de parole. Chaque ordinateur est configuré différemment dans le sens où les webcams ne sont pas toutes placées au même endroit par rapport à l'écran. Certains apprentis-enseignants utilisent leurs propres ordinateurs.

L'enregistrement par webcam nous offre plusieurs avantages souvent associés à des données écologiques ainsi que des avantages associées à des données contrôlées tout en réduisant le paradoxe de l'observateur.[18] Nous rejoignons Cicurel (2011 : 25) qui considère la classe de langue comme « un 'observatoire' qui permet de découvrir de quelle manière les données langagières sont maniées par les uns et les autres ». Tellier (2013 : 2) ajoute que « on peut, d'une manière générale, se demander si les données écologiques pures existent. En effet, à partir du moment où l'on entre dans une classe avec un magnétophone, une caméra ou seulement notre identité de chercheur, le naturel n'existe plus ». Une caméra placée dans une salle de classe est intrusive dans la mesure où elle risque de ne pas se faire oublier. En

[18] Selon le paradoxe de l'observateur (*observer's paradox*), le fait d'observer les participants est susceptible de changer le comportement de ceux-ci (Labov, 1973, 1976).

revanche, une webcam est moins intrusive car elle est là pour toute interaction visiophonique, que les interactants soient observés par un chercheur ou non. Le fait d'enregistrer les interactions et de les soumettre à des analyses scientifiques a-t-il un impact sur le comportement des interactants ? On ne saura répondre de manière définitive à cette question, mais Guichon (2013 : 105) postule que la webcam « est peu intrusive, les utilisateurs oubliant rapidement qu'un logiciel garde une trace de leur activité, et cette technique ne dénature pas l'activité des sujets comme cela est davantage le cas avec la présence d'une caméra ». Nous avons donc accès à des interactions très réalistes et naturelles, ce qui constitue un des avantages des corpus écologiques. Parallèlement, nous avons accès à des enregistrements qui ont été créés sous des conditions optimales, car tous les interactants portaient un micro-casque et les enregistrements viennent directement de la webcam.

Notre corpus d'interactions visiophoniques se distingue d'autres corpus de ce type recueillis récemment. Les interactions visiophoniques qui constituent le corpus de Develotte, Kern et Lamy (2011), par exemple, ont été provoquées spécifiquement pour la recherche scientifique[19]. Par contraste, les enregistrements multimodaux qui ont été recueillis pour ce corpus proviennent d'interactions ayant pour objectif principal l'enseignement/apprentissage du FLE. Par conséquent, les analyses menées dans le cadre du projet ISMAEL sont mieux adaptées à répondre à des questions didactiques et contribuer à la formation des enseignants.

4.9 Conclusion

Nous sommes arrivé à Lyon en 2014, un an après que les interactions ont eu lieu. La construction du corpus était déjà en cours avant notre arrivé, mais le corpus était incomplet. Ce qui manquait au corpus, c'était notamment la synchronisation des fichiers audiovisuels qui avaient été créés lors des interactions. Nous avons donc acquis certaines compétences techniques grâce à l'aide de la CCC du laboratoire ICAR. Notamment, nous avons appris à convertir et synchroniser des fichiers audiovisuels, à créer des « multiscopes », et à les exporter. Nous avons formé des étudiants en Master à utiliser le logiciel ELAN afin qu'un grand nombre de séances puissent être transcrites assez rapidement. Puis, nous avons mis ces interactions transcrites à disposition de plusieurs chercheurs ayant différentes questions de recherche. L'ouvrage de 2017 édité par Guichon et Tellier[20] est le fruit de ce travail

[19] La consigne fut « Discutez entre vous environ dix minutes sur le thème des relations hommes/femmes, femmes/femmes, ou hommes/hommes. Vous pouvez utiliser le chat écrit » (Develotte et al. 2011 : 12).

[20] Guichon, N., & Tellier, M. (Éd.). (2017). *Enseigner l'oral en ligne : une approche multimodale*. Paris: Didier.

collaboratif. Le rôle que nous avons joué à la construction du corpus a pris la plus grande partie de notre première année de thèse.

A cause de la nature écologique de notre corpus, nous avons fait le choix de partir de nos données pour consulter la littérature. Nous adoptons donc une démarche compréhensive (Demaizière & Narcy-Combes, 2007). Nous avons repéré des phénomènes dignes d'intérêt dans les données, comme les séquences d'explication lexicale et le caractère multimodal de celles-ci, avant de nous renseigner sur la recherche existante. Nous avons trouvé après-coup les conditions dans lesquelles les interactions se sont déroulées au lieu de les créer. Par conséquent, une démarche expérimentale n'était pas possible. Pour cela, il aurait fallu construire un cadre théorique et définir des variables contrôlables en amont des interactions. La démarche inductive que nous adoptons ici ne nous permet pas de formuler des hypothèses. Nos analyses seront donc qualitatives.

Chapitre 5 : Annotation des données et repérage des séquences d'explication lexicale

5.1 Introduction

Suite à la construction du corpus, un travail de réflexion et de transformation des données s'impose avant de pouvoir mener des analyses. En effet, la relation que nous entretenons avec nos données a une influence sur la façon dont nous repérons les phénomènes à étudier. Puis, la façon dont nous transformons nos données par des annotations oriente les analyses qui seront menées par la suite. Dans un premier temps, une réflexion personnelle et épistémologique sera proposée. Ensuite, nous présenterons la façon dont nous avons repéré les séquences d'explication lexicale faisant l'objet de nos analyses. Puis, nous aborderons l'annotation multimodale de nos données sur ELAN et sur Microsoft Excel ainsi que les choix méthodologiques associés.

5.2 Notre place par rapport à nos données

Comment nous positionnons-nous par rapport à nos données ? Cicurel (2011 : 8) rappelle que

> « La signification d'une action n'est pas la même pour le professeur qui impulse l'action, pour les apprenants en interaction avec lui et qui souvent ne découvrent son intention qu'au moment de l'échange, ou pour l'observateur qui n'est pas impliqué dans les échanges de la classe et ne participe pas aux 'jeux de miroir complexes' à l'œuvre dans l'interaction (Schütz, 1987) ».

Il nous importe donc de définir notre rôle et notre position par rapport à la communauté des enseignants en ligne au sens large, par rapport aux interactants que nous observons, et par rapport à nos données.

En CA[1], l'analyste essaie de comprendre ce qui se passe et quelle action est observée. Si celui-ci est membre du groupe qu'il observe, il est doté de savoirs qui l'aident à analyser les interactions. Être membre de la communauté que l'on observe permet de disposer d'informations ethnographiques (Ten Have, 1999). Dans le champ de la didactique des langues, Cicurel (2011 : 8) rejoint les chercheurs du champ de l'analyse conversationnelle :

> « [...] l'observateur n'est pas un *étranger* : il n'est pas extérieur aux « mœurs » observées, car il fait lui aussi partie de la communauté des enseignants et, à ce titre, entre l'analyse et les sujets agissants qui sont observés, existent des expériences partagées ».

[1] « Conversation analysis », ou analyse de la conversation.

Cette auteure (2011 : 20) ajoute qu'un observateur n'est pas un étranger car « il appartient à la communauté des enseignants » et « il partage une grande partie des préoccupations, des savoirs et des valeurs avec les acteurs observés ». Or, selon celle-ci il faut être étranger pour repérer les caractéristiques d'une culture éducative qui peuvent être ignorées par les membres internes au groupe car elles paraissent pour ceux-ci anodins, normaux, ordinaires, insignifiants. Notre place d'observateur pour chacune des trois catégories ci-après mérite d'être interrogée.

5.2.1 Notre place par rapport aux enseignants de langues étrangères

Nous avons deux ans d'expérience en tant qu'assistant de langue anglaise dans des écoles primaires de Pertuis et de Manosque (Académie d'Aix-Marseille). Pendant cette deuxième année, avons passé notre Diplôme d'Université « Parcours FLE » à l'Université d'Aix-Marseille où nous avons pris connaissance de l'histoire de l'enseignement du FLE en France et ses grandes lignes directrices méthodologiques et épistémologiques. Pendant notre Master FLE dans cette même université, nous avons observé et dialogué avec deux enseignantes chevronnées de FLE, ce qui nous a permis de voir de près les préoccupations et pratiques de celles-ci.

Notre expérience en tant qu'utilisateur de plateformes de visioconférence poste-à-poste doit être prise en compte car elle a un impact sur notre façon d'observer les interactants et d'interpréter leurs actes communicatifs. Nous avons expérimenté des plateformes de visioconférence poste-à-poste (Skype, Google Hangouts, etc.) en dehors du contexte d'apprentissage d'une langue avec des amis, des membres de la famille, et dans des contextes professionnels. Puis, nous avons animé trois séances d'anglais sur une plateforme d'enseignement (SpeakPlus[2]). Malgré la courte durée de cette expérience, celle-ci nous a néanmoins permis de comprendre certains phénomènes didactiques, émotionnels et pratiques auxquels se sont heurtés les interactants de notre étude. En particulier, nous avons ressenti la charge émotionnelle d'une première rencontre devant un écran avec un apprenant inconnu que l'on ne rencontrera probablement jamais en face à face. Nous avons construit les plannings des séances de manière similaire à celles construites par les apprentis-enseignants observés. Nous comprenons le besoin qu'ont les enseignants de terminer toutes les activités prévues et de respecter le temps accordé à celles-ci et la nécessité de veiller à ce que la personne en face

[2] SpeakPlus est le nom d'une startup lyonnaise. Leur plateforme est semblable à Visu.

à face à distance soit dans le confort requis par l'activité. Il faut aussi que les interactants prennent en compte les soucis techniques rencontrés dans l'utilisation de plateformes de visioconférence : décalages, coupures, mouvements saccadés, taux faible et variable du nombre d'images par seconde, problèmes d'audition, d'encadrement de la webcam, etc.

5.2.2 Notre place par rapport aux sujets de notre étude et à nos données

Ayant fait notre master recherche en didactique du FLE à l'Université d'Aix-Marseille, nous appartenons à la communauté des étudiants poursuivant ou ayant accompli leurs études supérieures de didactique du FLE en France, mais n'appartenant pas à la communauté des étudiants en master FLE professionnel et à celle des étudiants en master de l'Université Lumière Lyon 2. Pour pallier ce manque d'expérience, nous avons observé le cours de Master professionnel FLE de Nicolas Guichon en 2014 et nous avons co-animé avec celui-ci les mêmes cours l'année suivante. Cette expérience nous a permis de constater la collaboration étroite qui s'est mise en place entre l'Université Lumière Lyon 2 et Dublin City University, ainsi que les processus mis en œuvre par le formateur et les apprentis-enseignants pour concevoir et animer les séances hebdomadaires. Bien que deux modules de formation ne puissent jamais être identiques, nous travaillons avec l'hypothèse que les cours que nous avons observés en 2014 et co-animés en 2015 sont comparables à ceux de 2013, ceux à partir desquels notre corpus a été construit. Toutefois, on note la particularité suivante : en 2013 la plateforme utilisée était Visu, alors qu'en 2014, la plateforme était Adobe Connect, et SpeakPlus à partir de 2015. Nous estimons que le choix de la plateforme ne transforme pas profondément la nature des interactions, d'autant plus que nous adoptons une approche dite anthropo-centrée. Nos expériences nous accordent le statut de membre de la communauté que nous observons et nous autorisent à examiner les interactions en tant que tel.

Il faut remarquer que nous n'étions pas présent lors des interactions de 2013 et que nous ne donnions pas régulièrement de cours de FLE en ligne, ce qui peut aussi nous octroyer le statut d'étranger à la communauté alors constituée. C'est *a posteriori* que nous avons commencé à étudier les interactions qui constituent notre corpus. Cependant, notre participation à ce même cours pendant les deux années suivantes en tant qu'observateur et co-animateur nous a permis de comprendre le contexte qui avait été constitué en amont, même si celui-ci était un tant soit peu différent.

Nous avons donc un regard à la fois extérieur et intérieur sur nos données. Il convient d'aborder brièvement notre rapport à la langue qui a le double statut d'objet d'apprentissage et moyen de communication.

5.2.3 Notre rapport à la langue

Notre rapport à la langue française est similaire à celui entretenu par les apprenants de notre étude. D'origine des États-Unis d'Amérique, nous avons appris[3] tardivement le français comme langue étrangère. Au début du collège, dans le « sixth grade » (équivalent à la classe de sixième française), nous avons eu le choix entre l'espagnol et le français comme langue étrangère. Il n'était pas possible d'étudier les deux à la fois. L'espagnol étant beaucoup plus pratiqué dans le New Jersey que le français, la grande majorité des élèves choisit d'étudier l'espagnol. C'est donc délibérément que face à ce choix obligatoire nous avons commencé à étudier le français. Nous supposons que les apprenants dublinois adultes que nous observons apprennent volontairement le français pour des raisons professionnelles. Ce n'est que quatre ans plus tard, à l'âge de seize ans, que nous avons eu notre première professeure de français dont la langue maternelle était le français. Celle-ci parlait la langue de manière « authentique », sans accent anglophone, et avec plus de facilité que les autres enseignants que nous avions eus préalablement. Nous ignorons la qualité de l'enseignement dont ont profité les apprenants dublinois avant 2013, mais il se peut que certains d'entre eux se soient réjouis, comme nous l'avons fait à l'époque, de dialoguer avec un francophone natif, fût-il en présentiel ou à distance.

Notre apprentissage de la langue française nous permet de comprendre certaines difficultés langagières auxquelles font face les apprenants dublinois, d'autant plus que nous partageons la même langue maternelle, l'anglais. Par exemple, et pour ne mentionner que certains points liés au lexique[4], nous comprenons les difficultés qu'éprouvent Sean et Liam quand ils peinent à appréhender la différence entre *tu* et *vous*, différence que Samia et Victor essaient d'éclaircir, ainsi que les difficultés de prononciation liées à certains mots transparents, par exemple *gouvernement*[5], ou bien les difficultés de compréhension et de production induites par certains faux amis comme *magasin*[6], ou enfin les difficultés de genre qui n'existent pas en anglais. Pour ces différentes raisons, nous estimons avec modestie que

[3] Nous sommes d'avis que l'apprentissage de la langue n'est jamais terminé.
[4] Même si les connaissances lexicales, grammaticales et culturelles sont intimement liées.
[5] Chapitre 6
[6] Chapitre 7

nous sommes bien positionné pour reconnaître d'une part les préoccupations pédagogiques des apprentis-enseignants et d'autre part les difficultés langagières éprouvées par les apprenants.

Membre de la communauté des apprentis-enseignants, et ayant appris le français comme langue étrangère, nous avons un statut privilégié par rapport aux participants de notre étude, à nos données, et à la langue française.

5.3 Repérage des séquences d'explication lexicale

Après avoir fait le choix de nous concentrer sur le lexique, un grand travail de repérage s'est imposé.[7] Avant de démarrer ce travail, nous avons étudié les différentes façons de repérer des séquences destinées à l'analyse. Il existe plusieurs moyens d'identifier les séquences d'explication lexicale. Au sens large, il y a deux grandes approches pour repérer, étudier et définir un phénomène : une approche inductive et une approche déductive.

L'approche inductive consisterait à regarder le corpus sans avoir une définition d'une séquence d'explication lexicale au départ. Cette méthodologie, adoptée par l'analyse conversationnelle, ne prend pas en compte les théories ou définitions existantes car la CA a des intérêts émiques et non étiques[8] (Ten Have, 1999). Comme le dit Gülich (1991), « L'analyse conversationnelle vise à découvrir l'ordre dans le comportement des participants, non pas à l'imposer ». Les principes de la CA sont fondés sur l'expérience des membres au lieu d'être existants et appliqués. Ten Have (1999) s'appuie sur les travaux de Heritage (1988) pour résumer en trois étapes le processus que suivent les conversationnalistes :

1. En observant des données empiriques, on établit une régularité dans l'interaction ;

2. On décrit l'orientation normative des participants et on modifie progressivement la description du phénomène pour prendre en compte les cas déviants au fur et à mesure que ceux-ci surviennent ;

3. On fournit une description formelle du phénomène observé dans les étapes 1 et 2.

La première étape est plus inductive, traditionnelle et empirique, tandis que les étapes 2 et 3 sont plus analytiques. Cette distinction constitue les deux côtés de l'analyse conversationnelle. Lors de la première étape, on utilise beaucoup de données empiriques

[7] La *collectionabilité* (Ten Have, 1999) du phénomène est un autre critère qui justifie l'étude des séquences d'explication lexicale. Au lieu de mener des études cas isolées (*case studies*), ce choix de sujet nous permet de collectionner et d'étudier plusieurs séquences d'explication lexicale. Ce chapitre va donc traiter la méthode de collection.

[8] La recherche émique fait intervenir le point de vue intérieur des membres du groupe social. A l'inverse, la recherche étique se fait du point de vue de l'observateur, de l'extérieur.

tandis que lors de la seconde, on fait des études de cas précis. L'ordre dans lequel on réalise ces deux démarches n'a pas d'importance. Selon cet auteur, il faut les deux types de démarches pour arriver à des résultats intéressants.

Gülich (1991) plaide pour une approche ethnométhodologique similaire à l'approche utilisée par la CA telle que la décrit Ten Have (1999). L'approche ethnométhodologique, fondée par Garfinkel (1967), considère que les membres d'une communauté disposent d'un savoir commun et que les faits sociaux doivent être considérés comme des processus et non comme des choses. On n'analyse pas le monde tel qu'il est, mais tel qu'il est en train de se faire. Les normes sociales ne sont pas imposées sur les acteurs sociaux, mais sont continuellement créées par ceux-ci dans l'interaction. Il importe d'être membre du groupe qu'on observe afin d'être à même de comprendre les comportements des sujets qu'on observe.[9] Gülich (1991 : 338) reformule les trois étapes d'analyse proposées par Bergmann (1981, cité par Gülich, 1991) :

> « 1. découvrir des régularités dans le corpus, c'est-à-dire des structures verbales qui reviennent régulièrement dans certains contextes ; cela demande une observation précise ;
>
> 2. reconstruire le problème que les participants cherchent à résoudre à l'aide de ces structures ;
>
> 3. décrire la 'méthode' qui leur permet de résoudre le problème en question »

Le texte qui est étudié n'est pas un objet fini mais un processus. Ces trois étapes sont semblables à celles décrites par Ten Have (1999) dans le sens où on part des données pour arriver à une description du phénomène, et non l'inverse. Au lieu de décrire une régularité structurelle ou formelle, on décrit une procédure par laquelle les participants résolvent un problème communicatif.

A l'inverse, une approche déductive consisterait à trouver dans la littérature une définition d'une séquence d'explication lexicale et d'en trouver des exemples dans le corpus.

Nous nous proposons de donner quelques exemples concrets de repérage de séquences d'explication pour voir qu'une approche purement inductive n'est pas vraiment possible. Gülich (1991 : 357) prétend avoir appliqué une approche inductive, une méthode ethnométhodologique pour repérer les stratégies explicatives suivantes :

> « – la définition par synonyme ou par antonyme,
>
> – la comparaison,
>
> – la reformulation paraphrastique,
>
> – l'étymologie,

[9] Gülich remarque que l'analyse conversationnelle se rapproche de la linguistique car elle s'intéresse au langage comme objet d'étude. Mais l'analyse conversationnelle est d'abord sociologique et s'intéresse au langage car c'est une activité sociale. En didactique des langues, on s'intéresse à la langue car elle est à la fois objet d'apprentissage et outil de communication.

– la définition par 'genus proximum' et 'differentiae specificae',
– la traduction dans une autre langue,
– le recours à la situation,
– le recours aux connaissances encyclopédiques ».

Le problème avec cette classification est que l'auteure savait déjà qu'elle cherchait des stratégies d'explication. Cette catégorisation lui permet de décrire le phénomène recherché mais non pas de le définir.

Fasel Lauzon (2014), dans une démarche de l'analyse de la conversation, emploie une méthode assez courante pour repérer les séquences sur lesquelles elle travaille : elle identifie les séquences d'explication que les interlocuteurs considèrent comme telles. Elle recense donc les logonymes « expliquer » et « explication » dans ses transcriptions pour trouver les moments que les participants considèrent comme étant des explications. De Gaulmyn (1991 : 291) ne pense pas que la présence de commentaires métadiscursifs ou *métaexplicatifs* comme « expliquer » ou « comprendre » indique nécessairement la présence d'une séquence d'explication :

> « On sait d'autre part qu'énoncer le terme **expliquer** (même à la première personne du présent) ou le terme **comprendre** ne suffit ni à effectuer l'action, ni à signaler sa présence dans le discours (ce n'est pas un 'performatif' […]). Bien au contraire, la mention du mot peut justement dispenser d'effectuer l'action ou constituer une sorte d'anaphore, un rappel ou une annonce, un souhait ou un regret de l'explication absente. »

Il est cependant utile selon elle de recueillir des commentaires métadiscursifs qui peuvent témoigner de la présence, de l'absence ou de l'efficacité d'une explication. Pour De Gaulmyn, les commentaires méta-explicatifs sont des indices parmi d'autres qu'il convient de repérer.

Fasel Lauzon promeut une démarche inductive où on ne pose pas une définition du phénomène observé avant le repérage. Au lieu de définir en amont une séquence d'explication, elle repère les séquences intéressantes et les définit *a posteriori*. Or, le fait de recenser des logonymes comme « expliquer » peut correspondre à une définition d'une séquence d'explication lexicale. Selon nous, il est impossible de repérer un phénomène sans avoir une définition en amont, du moins une définition très générale. A l'instar de Miecznikowski (2005 : 41), qui a défini les séquences métalexicales comme « des séquences interactionnelles pendant lesquelles des locuteurs se focalisent réflexivement sur une unité lexicale du discours en train de se faire, pour traiter interactivement un problème qui se [pose à eux] à propos de cette unité », nous estimons qu'il est important de se fixer une définition en amont pour guider le repérage des séquences.

Nous avons défini une séquence d'explication lexicale comme un échange où une ou plusieurs facettes[10] de la connaissance d'un item lexical sont abordées, questionnées, traitées, discutées, ou expliquées par les participants. La séquence d'explication lexicale peut résulter d'un problème de communication, où un item lexical n'est pas compris par l'apprenant, ou bien elle peut résulter d'un problème de production où un apprenant a une lacune lexicale et cherche le *mot juste* (Grossmann, 2011). Nous considérons que les séquences d'explication lexicale monogérées où l'apprenant ne participe pas doivent être prises en compte et faire partie de notre corpus d'étude au même titre que les séquences d'explication lexicale qui sont dialogiques ou polygérées. Nous sommes d'accord avec Gülich (1991 : 334) qui argumente que :

> « Même si l'un des participants n'y contribue pas verbalement ou s'il s'agit du tour de parole d'un seul locuteur, le partenaire coopère à son accomplissement par le fait même que le locuteur s'adresse à lui ; par son existence même, l'interlocuteur coopère à la production discursive du locuteur, qui ne peut pas parler sans faire des hypothèses sur l'autre ».

En effet, quand l'apprenti-enseignant donne une explication non demandée, il fait des hypothèses sur les compétences langagières de l'apprenant. Ainsi, l'apprenant fait partie *ipso facto* de l'échange.

Nous avons visionné l'intégralité de notre corpus plusieurs fois, du début à la fin, afin de repérer toutes les séquences d'explication lexicale telles que nous les avons définies. Nous avons repéré 295 séquences d'explication lexicale. 131 séquences sont du premier type (problème de compréhension), soit 44%. 157 séquences sont du second type (problème de production), soit 53%. 7 séquences, soit 2%, sont les deux à la fois : un problème de compréhension et de production. Le tableau ci-dessous nous donne une vision globale des séquences.

	Nb.	%
Problème de compréhension	131	44,4
Problème de production	157	53,2
Les deux à la fois	7	2,4
Total	295	100

Figure 1 : Classification des séquences d'explication lexicale repérées

[10] Forme, sens, usage (cf. chapitre 2).

Les séquences d'explication lexicale se repartissent ainsi :

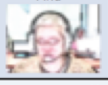
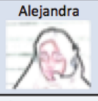
Apprentis-enseignants	Apprenants		Séance 1 15 octobre Les droits du travail	Séance 2 22 octobre Expériences professionnelles	Séance 3 12 novembre Se préparer pour un stage à Reims	Séance 4 19 novembre Gestion d'un projet	Séance 5 26 novembre Mettre en place un projet	Séance 6 3 décembre Entretien d'embauche
Adèle* 	Alannah 	Catriona* 	18	10	7	4	20	15
Emilie* 	Aiden 	Fiona 	11	11		8		8
Samia* 	Sean* 	Angela* 	11	8	1	6	12	
Etienne* 	Conor 	Sophie 				11		
Mélissa* 	Ana* 	Alejandra 	13	12	15	7	10	10
Séverine* 	Naomi 					21		
Victor* 	Liam 		12	5	9	12	8	

Figure 2 : Nombre de séquences d'explication lexicale repérées par séance transcrite

Dans chacune des 28 séances, le nombre de séquences d'explication lexicale varie entre une, dans le cas de Samia pour la troisième semaine, et 21, dans le cas de Séverine pour la quatrième semaine. Les cases vides représentent les séances qui n'ont pas été transcrites. Toute séance transcrite contient au moins une séquence d'explication lexicale, et le nombre moyen de séquences d'explication lexicale est de 10,5 par séance.

Après avoir repéré les 295 séquences, nous les avons délimitées et les avons divisées en phases suivant le modèle canonique évoqué dans le chapitre 3[11]. Parfois la délimitation était délicate, parce que le début et la fin de la séquence étaient souvent difficilement identifiables, ce qui rejoint le constat de De Gaulmyn (1991). Puis, nous avons comptabilisé, dans nos transcriptions multimodales, la façon dont les apprentis-enseignants se focalisent sur la connaissance des items lexicaux et la façon dont ils impliquent les apprenants dans les séquences d'explication. Le travail de repérage que nous avons effectué se résume comme suit.

[11] Ouverture, noyau, clôture.

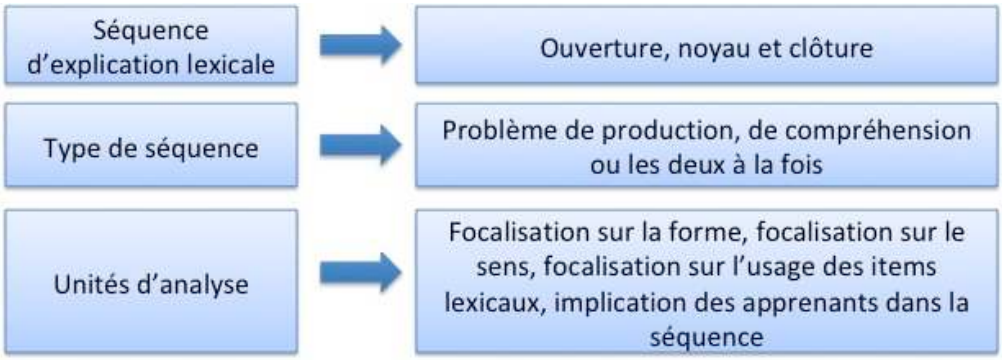


Figure 3 : Annotation des séquences d'explication lexicale

5.4 Annotation des phénomènes multimodaux

Aucune transcription n'est exhaustive, et chaque convention de transcription a un biais théorique et méthodologique (Ten Have, 1999). Quand on produit une transcription, on transforme un processus (l'interaction) en une chose (la transcription). Ensuite, le lecteur doit effectuer une transformation dans le sens inverse : il doit lire la transcription qui est une chose et reconstruire mentalement l'interaction qui est un processus. Nous avons transcrit nos données afin de rendre compte de la façon dont les apprentis-enseignants gèrent les séquences d'explication lexicale, à savoir la façon dont ils se focalisent sur la connaissance des items lexicaux et la façon dont ils impliquent les apprenants dans les séquences d'explication. Dans cette section, nous allons décrire notre façon de représenter nos données par la transcription et l'annotation. Le schéma ci-dessous synthétise les descriptions des annotations qui vont suivre.

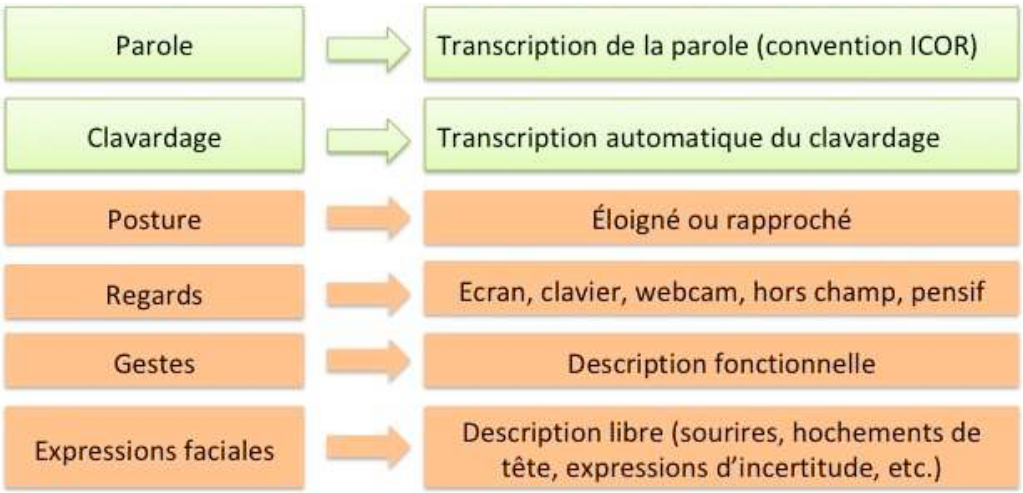


Figure 4 : Annotation des phénomènes multimodaux

5.4.1 Transcription de la parole

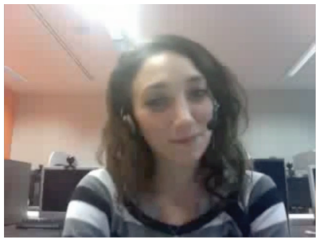
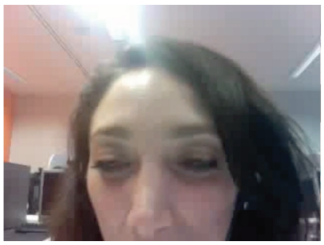
Nous avons mentionné dans le chapitre précédent que nous avons choisi la convention ICOR pour sa polyvalence. En effet, un schéma polyvalent comme celui-ci permet à plusieurs chercheurs avec différentes préoccupations scientifiques de travailler sur le même corpus. Les messages écrits envoyés par clavardage ont été copiés sans aucune modification. C'est en annotant des phénomènes multimodaux que nous avons personnalisé davantage nos annotations.

5.4.2 Annotation des ressources visuelles

Le nombre de phénomènes visuels à annoter est illimité. Cependant, il a fallu faire des choix afin de rendre les annotations lisibles et exploitables (Mondada, 2008). Nous avons essayé de prendre en compte le point de vue des membres grâce à notre expérience en tant qu'utilisateur de plateformes de visioconférence. Nous nous sommes posé les questions suivantes : quels indices audiovisuels sont perceptibles en visioconférence ? Qu'est-ce qui attire l'attention des interactants ? Nous avons aussi regardé les données pour saisir les éléments saillants. Nous avons donc décidé de diviser les phénomènes multimodaux en quatre grandes catégories, en annotant les postures, les expressions faciales, les regards et les gestes manuels. Nous les avons annotés sur ELAN (cf. chapitre précédent), et nous allons décrire notre façon de les annoter dans les sous-sections suivantes. Pour représenter nos annotations sur une page, nous avons élaboré un schéma à trois colonnes. Sur la colonne de gauche, nous avons mis la transcription de la parole orale en suivant la convention ICOR. Nous avons mis en caractères gras la partie de la transcription qui correspond temporellement à la capture d'écran que nous mettons dans colonne du milieu. Afin d'indiquer à quelle partie du tour de parole correspond l'image, nous avons ajouté sur ELAN une nouvelle piste d'annotation que nous avons nommée « capture d'écran » (voir Figure 5 en page 154). Ainsi, nous pouvons sélectionner à la milliseconde près le segment du tour de parole pendant lequel s'affiche une ou plusieurs images de la vidéo. Puisqu'il y a en moyenne deux ou trois images par seconde, chaque image correspond à 300-500ms de vidéo. Sur la colonne de droite, nous avons décrit les phénomènes dont la capture d'écran est l'illustration. Cette représentation à trois colonnes nous semble être un bon compromis entre exhaustivité et lisibilité (voir Exemple 1 ci-dessous).

5.4.2.1 Postures

Nous annotons les postures de manière beaucoup moins fine que Cohen (2017). D'un point de vue de la posture, un apprenti-enseignant peut être penché vers la webcam ou non. La webcam étant placée sur le même plan spatial que l'écran, l'apprenti-enseignant s'approche alors également de l'écran. Nous nous intéressons à la fonction communicative remplie par un rapprochement de l'écran lors d'une séquence d'explication lexicale. Par exemple, lorsqu'un étudiant fait une erreur de production liée au lexique, et qu'une séquence d'explication lexicale s'impose pour trouver le mot juste, un apprenti-enseignant peut se rapprocher de l'écran pour diriger l'attention de l'apprenant vers l'item langagier et ouvrir un espace de collaboration dans lequel le mot juste peut être trouvé (cf. chapitre 8). C'est le placement séquentiel de ce mouvement qui nous intéresse plutôt que le degré exact du cadrage de la webcam sur une échelle à cinq niveaux (Cohen, 2017)[12]. Nous allons illustrer à l'aide d'un exemple notre façon d'annoter un rapprochement de l'écran.

Transcription de la parole	**Image du participant**	**Phénomènes observés**
1 ALN **mon euh mon père/ a pu travailler/ à (0,5) à chez moi\ (2,0)**		Adèle est éloignée de l'écran
2 ADL ton père quoi/ **(1,1)**		Adèle se rapproche à l'écran

Exemple 1 : Rapprochement à l'écran

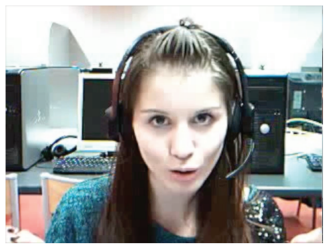
Dans cet exemple, Adèle se rapproche à l'écran quand elle ne comprend pas la parole d'Alannah. Elle va par la suite donner la forme correcte de « travailler de chez soi ».

5.4.2.2 Regards

La mauvaise résolution de la vidéo, ainsi que la distance entre la webcam et le visage de l'apprenti-enseignant, rendent la direction du regard difficilement perceptible. Cependant,

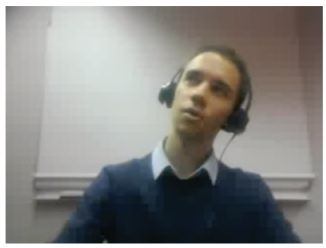
[12] Cohen (2017 : 221) distingue cinq niveaux du cadrage de la webcam : *très gros plan, gros plan visage, gros plan tête et épaules, plan tête et épaules,* et *plan américain*. Elle définit cette grille afin de mesurer l'évolution de l'utilisation de la webcam au fil du semestre

nous pouvons distinguer certaines directions générales, à savoir un regard vers l'écran, vers le clavier, vers la webcam, ou hors-champ. A ces quatre directions nous ajoutons la catégorie « regard pensif » pour sa valeur communicative. Quand un apprenti-enseignant produit un regard pensif, on peut faire l'hypothèse qu'il est en train de réfléchir. A l'aide de deux exemples, nous allons montrer notre façon d'annoter la direction des regards. Dans l'exemple suivant, Mélissa vérifie sa propre compréhension de ce que son apprenante Ana cherche à dire. Avant de donner le mot « promotion », Mélissa donne un exemple et regarde directement vers la caméra pour solliciter une réponse d'Ana.

Transcription de la parole	**Image du participant**	**Phénomènes observés**
1 MEL et si vous êtes dix ça va coûter (1,1) sept **euros**		Regard caméra

Exemple 2 : Regard caméra

Dans l'exemple suivant, Victor et Liam cherchent un mot juste. Avec un regard pensif, Victor propose le mot « entraînement ». Le regard pensif, ou « thinking face » anime la recherche lexicale (Goodwin & Goodwin, 1986 : 57).

Transcription de la parole	**Image du participant**	**Phénomènes observés**
1 VIC plus de: hum **(0,3)** 2 LIM **plus de** **(0,3)** 3 VIC **d'entraînement peut** être/		Regard pensif

Exemple 3 : Regard pensif

Bien que nous ne disposions pas d'une machine qui enregistre avec précision la direction des regards des interactants[13], nous pouvons tout de même annoter certains regards saillants comme le montrent ces deux exemples.

[13] Le Laboratoire Parole et Langage d'Aix-en-Provence dispose d'une machine qui enregistre avec précision tous les mouvements des yeux afin d'analyser ce que regarde un sujet sur l'écran. Une telle machine permettrait de savoir avec exactitude quelle partie de l'écran est regardée par les participants de notre étude.

5.4.2.2.1 Un vocabulaire contrôlé

Puisque nous avons défini un nombre limité de regards possibles, nous avons créé sur ELAN ce qu'on appelle un *vocabulaire contrôlé*. Cela veut dire que nous avons préenregistré tous les types de regards possible. L'annotation est ainsi facilitée car il suffit de sélectionner la direction du regard parmi une liste qui s'affiche sur une liste déroulante. Cela évite aussi la possibilité de faire des fautes de frappe, ce qui peut nuire à la capacité d'effectuer des recherches ou comptages ultérieurement. La capture d'écran ci-dessous montre le vocabulaire contrôlé pour les regards.

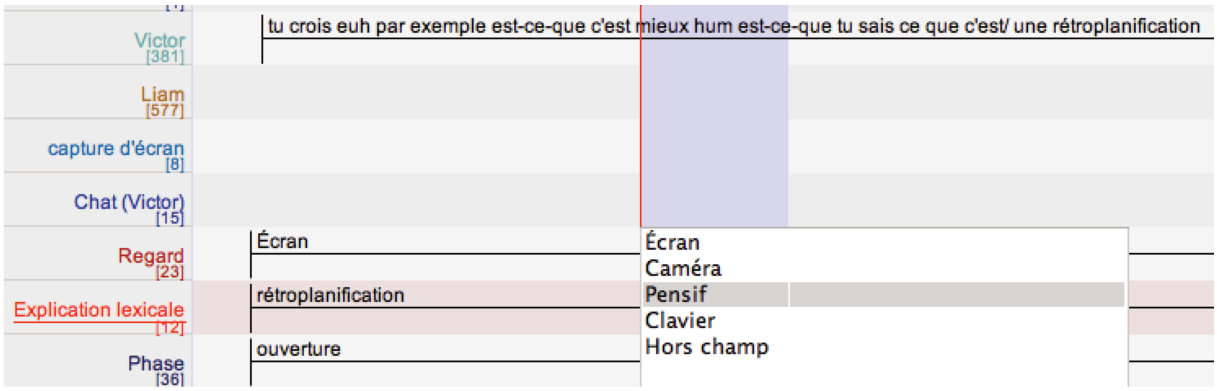
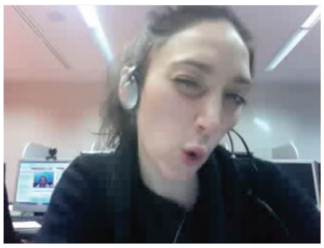
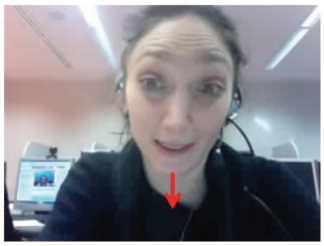


Figure 5 : Vocabulaire contrôlé

Nous avons créé des vocabulaires contrôlés pour annoter les postures (rapproché ou non) et les regards (écran, caméra, pensif, clavier, hors champ). Pour les autres phénomènes, nous les avons annotés librement à cause de leur complexité.

5.4.2.3 Expressions faciales

Contrairement aux gestes, dont la visibilité est incertaine, les expressions faciales sont *a priori* toutes visibles à condition que le visage de l'apprenti-enseignant reste devant la webcam. Six apprentis-enseignants sur sept se montrent visibles en permanence. Nous décrivons librement les expressions faciales plutôt que de les catégoriser, même si certaines expressions faciales sont récurrentes et symboliques par convention comme des expressions d'incertitude que fait un apprenti-enseignant lorsqu'il ne comprend pas son apprenant ou quand il vérifie la compréhension d'un item lexical. La transcription ci-dessous exemplifie la façon dont nous annotons les expressions faciales et les hochements de tête.

Transcription de la parole	Image du participant	Phénomènes observés
1 ALN on euh on est/ (0,2) **on a/ (0,5) honnête/ (0,2)**		Visage d'incertitude
2 ADL **honnête**		Haussement de sourcils suivi d'un hochement de tête

Exemple 4 : Annotation des expressions faciales et hochements de tête

Alannah cherche à prononcer le mot « honnête ». Lors de ses deux premières tentatives, Adèle réagit avec un visage d'incertitude qui lui donne du feedback. Quand Alannah trouve la bonne prononciation, Adèle le valide en répétant le mot avec un haussement de sourcils et un hochement de tête. De même que nous ajoutons des flèches pour représenter le mouvement des mains, nous ajoutons des flèches pour représenter les hochements de tête.

Les regards et les expressions faciales ont été distingués pour faciliter l'annotation, mais ils sont tous les deux considérés comme étant des ressources visuelles. La division entre les expressions faciales et les regards est moins importante que les fonctions communicatives remplies par ces ressources. Un « regard pensif » peut par exemple être considéré comme une expression faciale au même titre qu'un « regard d'incertitude ». La distinction a été faite pour faciliter l'annotation et non pas pour catégoriser l'acte communicatif.

5.4.2.4 Gestes

Qu'est-ce qui peut être considéré comme un geste ? McNeill (1992 : 1) définit les gestes comme « les mouvements des mains et des bras que l'on voit quand des personnes parlent. »[14]. Or certains mouvements, comme fumer, sont extracommunicatifs et ne comptent pas comme des gestes (Kendon, 2004; McNeill, 1992; Tellier, 2009a)[15]. Pour Kendon, l'intention communicative du geste est primordiale. Quand on regarde un enregistrement

[14] « the movements of the hands and arms that we see when people talk. »

[15] Cependant, Goodwin (1981) a montré que certaines actions comme fumer peuvent être instrumentales et communicatives. Par exemple, un interlocuteur peut commencer à fumer pour montrer qu'il ne s'intéresse plus à la conversation.

vidéo d'un sujet en train de parler, on peut dire intuitivement quels gestes sont hétéro-centrés. En effet, l'identification de gestes hétéro-centrés est facilitée parce que la plupart des gestes sont invisibles et hors champ de la webcam (Guichon & Wigham, 2016).

L'annotation des gestes produits en visioconférence n'est pas la même chose que l'annotation des gestes produits dans une salle de classe ou dans une chambre sourde[16]. Comme évoqué dans le chapitre 3, la production de gestes en visioconférence est particulière car la webcam a un effet sur la visibilité du corps de l'interlocuteur. Même si on voit les mêmes gestes en face à face et sur un écran, il peut y avoir des différences au niveau de la perception des gestes. Beattie & Shovelton (1999) ont étudié l'effet des gestes vus sur un écran et vus en face à face. Ils ont trouvé que les gestes iconiques vus en face à face étaient plus efficaces et donnaient plus d'informations que les gestes iconiques vus sur un écran. En se demandant pourquoi le face à face a été plus efficace que la vidéo, ils postulent que la qualité de l'image, la présence d'indices sociaux (comme le fait de regarder ses propres gestes), et la taille de l'écran pourraient jouer un rôle. En visioconférence, la taille de l'écran et la qualité de la vidéo ont certainement un effet sur la production et la réception de gestes.

5.4.2.4.1 Espace gestuel

Puisque la visibilité de l'interlocuteur est réduite en visioconférence, l'annotation des gestes doit s'adapter. Dans cette section, nous allons expliquer pourquoi l'espace gestuel, qui est une notion incluse dans un grand nombre d'études gestuelles, ne sera pas pris en compte dans nos annotations. McNeill (1992) a étudié des sujets assis en train de raconter des histoires de bandes dessinées qu'ils ont visionnées. Les corps des sujets étaient pleinement visibles, et l'espace gestuel était représenté ainsi :

[16] Une chambre sourde, autrement appelée chambre anéchoïque, absorbe les ondes sonores et permet d'enregistrer la parole sans aucune nuisance sonore.

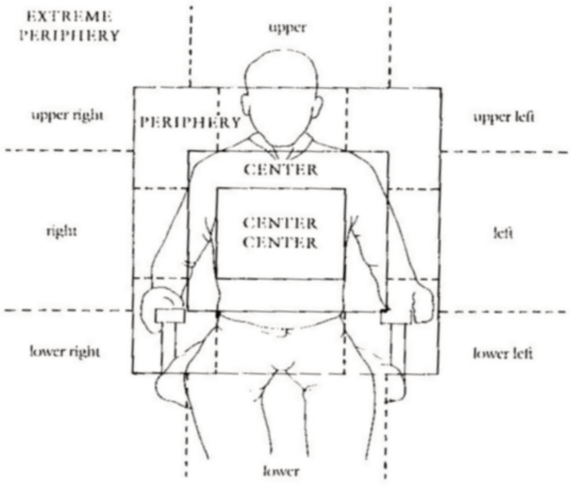
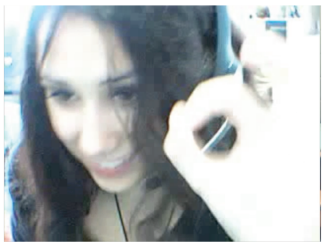


Figure 3.1. Drawing of the typical gesture space of an adult speaker.

Figure 6 : Espace gestuel de McNeill (1992 : 89)

L'espace gestuel tel qu'il a été défini par McNeill a des conséquences sur la façon d'étudier les gestes. Notamment, une distinction peut être faite entre un geste produit dans le centre et un geste produit dans la périphérie de l'espace gestuel. McNeill se sert de sa notion d'espace gestuel pour dire que différentes cultures utilisent l'espace gestuel à leur façon. Il s'en sert également pour distinguer les différentes dimensions gestuelles qu'il décrit. En général, les iconiques occupent l'espace « centre-centre », les métaphoriques occupent le bas de l'espace centre, les déictiques s'étendent à la périphérie, mais les battements dans tout l'espace, leurs emplacements étant différents selon le locuteur. Pour McNeill, le fait que les gestes occupent des zones différentes justifie davantage leur division en quatre dimensions : iconiques, métaphoriques, déictiques, et battements.

En visioconférence, l'espace gestuel est nettement réduit, ce qui a un impact sur la visibilité des gestes, comme le notent Cosnier & Develotte (2011) et De Chanay (2011). Nous allons illustrer ce phénomène à l'aide d'un exemple. Dans la transcription suivante, Emilie produit un geste emblématique voulant dire « un zéro » quand elle explique le sens de l'expression « arrondir ses fins de mois » :

Transcription de la parole	Image du participant	Ph. observés
1 EMI on est sensé rajouter **UN ZÉRO/ (.) d'accord donc** avoir une paye\ un peu plus grosse (.) un peu plus importante\		Geste emblématique représentant un zéro

Exemple 5 : Geste produit dans le champ de la webcam

En utilisant l'espace gestuel de McNeill reproduit ci-dessus, dans quelle zone se trouverait le geste d'Emilie ? On peut dire qu'il est *a priori* au centre, car on le voit au centre de l'écran. On peut dire aussi que celui-ci est dans la périphérie, car il se produit au-dessus de la tête. Puisque la main d'Emilie est près de la webcam, l'étendue de l'espace gestuel n'est que de quelques centimètres. De plus, Emilie décale sa tête pour faire place à son geste, ce qui change l'espace gestuel perceptible. Il semble qu'Emilie, comme nous, considère que le geste doit simplement être « dans le champ ». L'espace gestuel de McNeill semble inadapté à l'analyse de gestes en visioconférence. Au lieu de dire que le geste est produit dans le centre, dans la périphérie, à gauche, ou à droite du champ de la webcam, il semble plus pertinent de dire que le geste est produit ou « dans le champ », ou « hors champ ». Guichon & Wigham (2016), en utilisant une caméra extérieure filmant les apprentis-enseignants en train de communiquer sur Visu, ont trouvé que la plupart des gestes étaient produits hors champ de la webcam. Cela fait intervenir le point de vue de l'analyste (étique) plutôt que le point de vue des membres (émique). Nous, comme les interactants, ne prenons en compte que les gestes qui sont produits dans le champ. Certains gestes sont produits près de la webcam, comme celui produit par Emilie dans l'exemple ci-dessus, ce qui les rend plus saillants.

5.4.2.4.2 Annotation des mouvements

Les gestes sont souvent annotés par des *unités gestuelles* (Kendon, 2004; McNeill, 1992). Une unité gestuelle est composée d'une phase préparatoire, un *stroke*[17], et une phase de rétractation. En partant de la position de repos, les mains font ce que Kendon (2004 : 11) appelle un « *movement excursion* ». Pendant la phase préparatoire, les mains partent de la position de repos vers le lieu où le geste doit s'exprimer et créer du sens pendant le *stroke*. Avant le *stroke*, les mains peuvent, d'une manière non obligatoire, s'arrêter pendant le *pre-stroke hold*. Ensuite, d'une manière non obligatoire, le geste peut être maintenu par un effort musculaire. On parle alors d'un *post-stroke hold*[18]. Pendant la phase de rétractation, les mains reviennent à la position de repos. Le *noyau* est donc composé du *stroke* et du *post-stroke hold*. Une *phase gestuelle* est composée de la phase préparatoire, du *stroke* et *du post-stroke hold* (le noyau). Il peut y avoir plusieurs phases gestuelles dans une seule unité gestuelle.

Afin d'annoter des gestes de cette façon, deux conditions au moins doivent être remplies. On doit voir d'abord la position de repos des mains afin de pouvoir délimiter les unités gestuelles et ensuite, l'enregistrement vidéo doit être assez fluide pour pouvoir

[17] En anglais, la traduction littérale est « caresse de la main ».
[18] En anglais, la traduction littérale est « tenir la main (en l'air) après le stroke ».

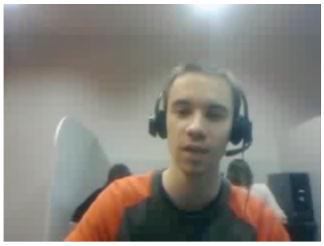
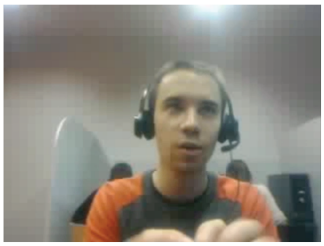
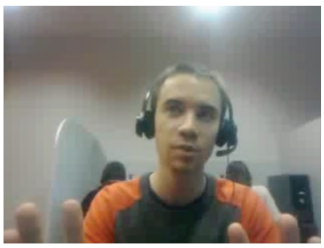
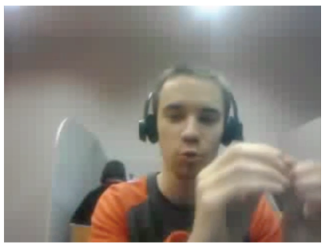
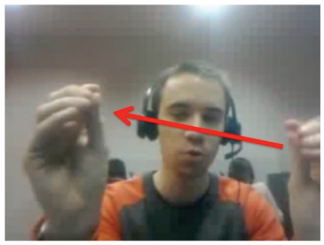
apercevoir les noyaux et délimiter les phases gestuelles. Les enregistrements vidéo dont nous disposons ne remplissent malheureusement pas ces deux conditions. Concernant la position de repos, on ne voit que la tête et les épaules des interactants la plupart du temps. Quand un geste est produit, on ne voit jamais la position de repos, à moins que les mains ne partent de la tête de l'apprenti-enseignant. Quand un geste est produit totalement ou partiellement en dehors du champ de la webcam, soit il n'y a rien qui apparaît à l'image, soit on ne voit que le bout des doigts et le geste est non identifiable. Si le geste est produit dans le champ de la webcam de manière rapide, l'image qui apparaît aux interactants est floue. Enfin, même si le noyau du geste est visible, on peut l'identifier mais on ne reconnaît pas toujours le nombre de gestes en raison du caractère saccadé des enregistrements vidéo. La qualité de la vidéo est dégradée par rapport à une vidéo prise par une caméra classique en raison du débit de la connexion Internet. En effet, le nombre d'images par seconde est faible, et la résolution de la vidéo est très basse, ce qui fait que les mouvements que l'on voit sont extrêmement saccadés. Il est donc difficile pour les interactants d'observer le mouvement des mains, et ainsi de distinguer plusieurs phases gestuelles.

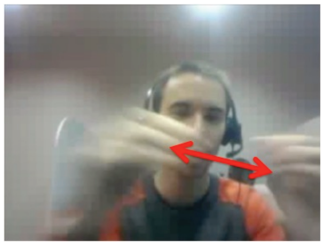
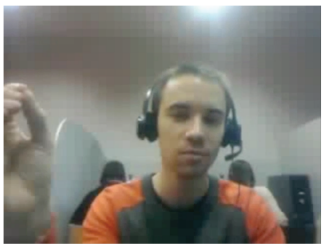
Puisque les difficultés mentionnées ci-dessus sont des problèmes pratiques pour les participants avant d'être des problèmes pour l'annotateur, nous faisons l'hypothèse que les participants adaptent leurs gestes en les simplifiant, en les caricaturant, en les maintenant plus longtemps, en les rendant plus visibles, afin de faire face aux contraintes de résolution et de fluidité. Dans le cadre de notre mémoire de Master 2, nous avions fait l'hypothèse que les apprentis-enseignants allaient produire plus de gestes lors des séquences d'incompréhension. Or cette hypothèse n'a pas été validée. Cependant, il semblerait que qualitativement, les apprentis-enseignants portent plus d'attention à leurs gestes quand ils savent que l'apprenant est en difficulté. Nous allons illustrer ce phénomène à l'aide d'un exemple qui nous donne aussi l'occasion de décrire notre façon d'annoter les mouvements des gestes.

La partie du tour de parole qui correspond à l'image est écrite en caractères gras. Pour annoter le mouvement, nous mettons une flèche rouge par mouvement détecté lors du noyau de la phrase gestuelle. Si un geste est maintenu, nous ne mettons pas de flèche. Malheureusement, le nombre d'images par seconde est tellement faible qu'il arrive qu'une flèche rouge corresponde non pas à un mouvement discret mais à une image[19].

[19] Par exemple, si un interlocuteur fait un mouvement fluide avec sa main pendant une seconde, on aura l'impression de voir trois mouvements discrets et saccadés car nous avons trois images par seconde.

Dans l'exemple suivant, Victor parle des compétences requises pour un stage que son apprenant aura à effectuer en France.

Transcription de la parole	Image du participant	Phénomènes observés
1 VIC et et est-ce que ce projet pourrait euh:/ c- le projet dont tu m`mparles/ (0,4) est-ce que tu pourrais l'utiliser/ pour euh vendre euh ta candidature/ (0,3) pour le stage\ (0,3) ou y a pas de **lien** (3,3) 2 LIM pardon/ euh répétez s'il vous pl[aît/]		Aucun geste n'est produit
3 VIC [est-ce] que tu arrive**rais à faire un lien/** entre ton expérience (0,3) de: que tu m`parles maintenant/ (0,3) et: le: stage (0,6) que: que j`t'ai envoyé/ l'annonce de stage (0,7)		Regard pensif et geste métaphorique peu visible représentant le concept d'un lien
est-ce qu'il y a des liens possibles/ ou c'est **du c'est c'est pas du** tout pareil/ (2,4) 4 LIM hum: (0,4) qu'est-ce que c-qu'est/ (0,6) qu'est-ce que c'est euh des liens/ (1,1) 5 VIC qu'est-ce que c'est quoi/ (0,6) 6 LIM des liens (0,3)		Geste emblématique peu visible et regard pensif
7 VIC ah des liens/ c'est hum: (0,9) euh: une relation en- entre **une chose**		Début d'un geste métaphorique bien encadré.
et une autre euh on dit		La main droite se sépare de la main gauche pour illustre le concept d'un lien

qu'il y a un lien		Victor refait son geste illustrant un lien. Les deux mains sont mis ensemble au milieu de l'écran et se séparent.
(0,5) 8 LIM **ah** oui oui oui		Victor maintient la fin de son geste pendant 800ms

Exemple 6 : Annotation d'une unité gestuelle

Victor utilise le mot « lien » à la ligne 1 sans faire de gestes. Lors de ce tour de parole, il n'y a pas une focalisation sur la langue. Liam indique à la ligne 2 qu'il ne comprend pas et demande une répétition sans préciser l'élément qui a provoqué l'incompréhension. Victor, en ignorant quel élément de son tour précédent a provoqué l'incompréhension, reformule sa question en faisant des gestes légèrement visibles qui illustrent le mot « lien » et l'expression « c'est pas du tout pareil » (ligne 3). Puis, aux lignes 4, et 6, Liam précise que c'est le mot « lien » qu'il ne comprend pas. A la ligne 7, Victor est conscient que le mot « lien » est incompris et fait des gestes amples et bien visibles lors du noyau explicatif. Il fait deux gestes métaphoriques illustrant le concept de « lien », et maintient son geste pendant 800ms jusqu'au début du tour de parole suivant de Liam (ligne 8). Victor semble considérer que ces gestes sont une ressource utile lors des séquences d'explication lexicale, et il semble être conscient de leur visibilité. On assiste à un *recipient design* (Sacks, Schegloff, & Jefferson, 1974) multimodal.

Victor est conscient du fait que le geste doit être encadré car lors d'un débriefing, il dit que la technologie est « un obstacle », et que :

> « La technologie, pour moi, depuis le début je la prends comme… C'est pas, c'est pas naturel, c'est pas authentique. Après, c'est pas, c'est pas désastreux, ça fait partie de cette expérience et puis c'est tout ! C'est pas un problème. Mais, oui je sens que dans la façon de s'exprimer ou dans la gestuelle, quand tu es obligé de t'aligner sur la caméra pour faire ton geste… »

Tout porte à croire que cet apprenti-enseignant est conscient du fait que les gestes nécessitent une certaine conscientisation afin d'être rendus visibles en visioconférence.

5.4.2.4.3 Classification des gestes

Une fois les gestes repérés, leur classification s'impose. Il existe plusieurs manières de classer les gestes coverbaux, et ce choix doit servir à l'analyse qu'en font les participants et *a posteriori* l'analyste. Comme le rappelle Kendon (2004), toute catégorisation n'est qu'un dispositif temporaire pour des études locales. McNeill (1992) est l'auteur d'un des schémas d'annotation les plus utilisés par les gestualistes. Il divise les gestes en quatre dimensions : les iconiques, les métaphoriques, les déictiques et les battements. Dans le cadre de notre Master 2, et dans Holt, Tellier & Guichon (2015) nous avons annoté quatre séances de la première semaine en classant les gestes ainsi. Nous avons donc créé un vocabulaire contrôlé sur ELAN avec ces cinq catégories de gestes.

La classification de gestes par dimension n'a pas beaucoup d'intérêt pour une étude sur l'explication des items lexicaux en visioconférence. Ce qui nous intéresse, ce sont les fonctions des gestes lors des séquences d'explication lexicale. Certains chercheurs comme Tellier (2008a) et Sime (2006, 2008) classifient les gestes pédagogiques par fonctions.

Sime (2006, 2008) s'intéresse aux fonctions pédagogiques des gestes des enseignants de l'anglais langue étrangère du point de vue des apprenants qui les observent *a posteriori* sur un enregistrement vidéo. Elle a trouvé que les apprenants faisaient attention aux gestes quand ceux-ci avaient un effet sur l'apprentissage. Notamment, les apprenants prêtaient attention aux gestes en situation de détresse langagière quand la compréhension orale était difficile. Les apprenants décodaient avec facilité le langage corporel des enseignants, et la plupart du temps, quand ils étaient confrontés au même geste, les apprenants étaient d'accord entre eux sur le sens du geste.

Sime a trouvé que les apprenants ont attribué trois fonctions aux gestes produits par les enseignants : une fonction cognitive, une fonction émotionnelle/affective, et une fonction organisationnelle. Concernant la fonction émotionnelle, une posture relâchée, un regard maintenu, et la présence de beaucoup de gestes et d'expressions faciales servent à encourager l'apprenant et à lui donner du feedback. Concernant la fonction organisationnelle, l'usage d'emblèmes et de déictiques sert à gérer les tâches et les tours de parole.

Les fonctions cognitives des gestes accomplissent trois sous-fonctions : (1) améliorer la compréhension, (2) faciliter le processus d'apprentissage, et (3) indiquer une réaction à l'output de l'apprenant. D'après les interviews avec les apprenants, les gestes produits par l'enseignant peuvent selon eux améliorer la compréhension « en illustrant des mots ou des idées, en soulignant des mots importants, ou en marquant des contrastes et des

comparaisons » (Sime, 2008 : 275, notre traduction)[20]. Les enseignants font des gestes pour représenter visuellement certains mots qu'ils estiment être difficiles à comprendre. Ces gestes aident les apprenants à formuler ou confirmer des hypothèses selon le degré de connaissance de l'item lexical. Puis, les gestes qui facilitent le processus d'apprentissage servent parfois à orienter l'attention de l'apprenant vers un élément langagier important comme un item lexical ou un point de grammaire. Ils servent aussi à inciter l'apprenant à trouver le mot juste. Enfin, les enseignants utilisent leurs corps, surtout les hochements de tête, pour donner du feedback et valider l'output de l'apprenant.

Tellier (2008a) nomme aussi trois fonctions des gestes pédagogiques : informer, animer, et évaluer. L'auteure explique qu'il y a trois types de gestes pour informer : les gestes d'informations grammaticales, les gestes d'informations lexicales, et les gestes d'informations phonologiques et phonétiques. On retrouve ces gestes d'information dans les séquences d'explication lexicale quand un apprenti-enseignant explique une facette de la connaissance d'un item lexical. Selon la conception que nous avons adoptée dans le chapitre 2 de la connaissance d'un item lexical par facettes (forme, sens et usage), les trois types de gestes d'information décrits par Tellier peuvent tous être lexicaux. Selon Singleton (1994 : 3), « on se rend compte que le lexique et la grammaire s'interpénètrent au point que leur séparation pose un vrai problème », donc on n'opère pas une distinction entre grammaire et lexique, ce qui fait que les gestes d'information grammaticale comportent donc aussi une information lexicale. Concernant les gestes d'information phonologique, connaître un item lexical suppose que l'on connaisse ses formes orale et écrite (la forme que représente le geste) et que l'on en ait une bonne représentation phonologique associée. Un geste d'information phonologique comporte donc aussi une information lexicale. Pour les gestes pédagogiques servant à informer, nous nous intéressons simplement aux gestes de l'information lexicale, quelle que soit la facette de la connaissance abordée.

Tellier (2008a : 4) explique que le geste d'animation

> « englobe à la fois les gestes de gestion de classe (changement d'activité, démarrage et clôture d'activité, placement des apprenants/ du matériel, punir/gronder/faire taire, donner des consignes) et de la gestion des interactions et de la participation (réguler les débits/le volume sonore, faire répéter, étayer, interroger, donner la parole) ».

[20] "Some gestures were identified by the learners as enhancing comprehension of the teacher's message either by illustrating words or ideas, emphasizing important words or marking contrasts and comparisons" (Sime, 2008 : 265)

On retrouve ces gestes notamment dans les phases d'ouverture et de clôture des séquences d'explication lexicale et lors des passages entre les phases. On les retrouve aussi lorsque les apprentis-enseignants gèrent les tours de paroles.

Enfin, les gestes d'évaluation servent à « *féliciter*, *approuver* et *signaler une erreur* » (Tellier, 2008a : 4). On retrouve ces gestes suite à une erreur lexicale d'un apprenant, ce qui déclenche une séquence d'explication lexicale. On les retrouve également lors de la clôture d'une séquence d'explication lexicale où l'apprenant félicite l'*uptake* de l'item lexical ou de sa connaissance.

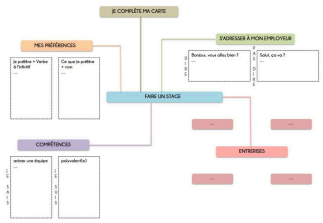
La classification de gestes par fonction pédagogique proposée par les deux auteures citées ci-dessus nous semble plus adapté à l'analyse des séquences d'explication lexicale qu'une catégorisation par dimension à la McNeill. Pour nos analyses, nous annotons les gestes suivants :

- Les gestes d'information qui servent à expliquer une facette de la connaissance d'un item lexical (forme, sens, et usage) ;
- Les gestes d'animation qui servent à gérer le déroulement d'une séquence d'explication lexicale, le passage entre les phases ou la gestion des tours de parole ;
- Les gestes d'évaluation qui servent à donner du feedback aux apprenants sur leur production langagière avant, pendant ou après une séquence d'explication lexicale et qui servent à encourager les apprenants à les impliquer affectivement.

Nous analyserons dans les chapitres 6 et 7 les gestes qui sont employés par les apprentis-enseignants afin de se focaliser sur la forme et sur le sens des items lexicaux. Dans le chapitre 8, nous analyserons les gestes qui sont produits par les apprentis-enseignants pour impliquer les apprenants dans les séquences d'explication lexicale.

5.4.2.5 Documents envoyés

Nous n'avons pas d'accès direct aux images telles qu'elles s'affichaient dans Visu, mais certaines images ont été envoyées par des liens. Pour les images auxquelles nous avons accès grâce aux liens, nous les représentons dans la deuxième colonne là où apparaît normalement les images des participants. La transcription suivant exemplifie notre façon de reproduire les documents qui ont été envoyés par Visu.

Transcription de la parole	Image du participant	Phénomènes observés
1 VIC je vais t'envoyer la carte mentale/ (0,2) comme ça peut-être que tu pourras euh voir/ de quoi ça a l'air\ **<((image))>** dans le chat		Victor envoie une image d'une carte mentale

Exemple 7 : Annotation des images envoyées

5.4.2.6 Synthèse : annotation multimodale

Nous reconnaissons qu'il est impossible de tout transcrire, et qu'aucune transcription n'est neutre. C'est donc par choix, sans aucune prétention d'exhaustivité, que nous avons divisé l'annotation des phénomènes multimodaux (hormis la transcription de la parole et du clavardage) en quatre catégories : les postures, les regards, les gestes, et les expressions faciales. Quand les photos et documents envoyés sont disponibles, nous les incluons. Pour faciliter la transcription des postures et regards, nous avons créé des vocabulaires contrôlés sur ELAN. Les gestes et expressions faciales ont été annotées librement pour prendre en compte leur complexité. Nous nous intéressons aux fonctions des gestes lors des séquences d'explication lexicale plutôt que de leur type. Dans les prochains chapitres, nous analyserons comment les ressources que nous avons annotées sont employées par les apprentis-enseignants pour gérer les explications lexicales.

Chaque représentation des données implique une transformation ainsi que des choix. Nous avons opté pour une représentation des interactions à trois colonnes : une colonne à gauche pour la parole orale et écrite, une colonne au centre pour les ressources visuelles, et une colonne à droite pour une description des phénomènes représentés dans la deuxième colonne. Tandis que la transcription de la parole dans la première colonne est quasiment exhaustive, la deuxième colonne est sélective. Nous ne sélectionnons que des images qui selon nous contribuent à la gestion des séquences d'explication lexicale et qui sont saillantes. Nous reconnaissons la subjectivité de ces annotations, est c'est la raison pour laquelle nous avons ajouté la troisième colonne afin d'expliquer au lecteur ce que l'on observe. Nous pourrions parler d'*intersubjectivité* car nous sélectionnons les phénomènes à annoter en tant que membre de la communauté que nous observons. Notre choix de limiter nos tableaux d'annotation à trois colonnes est lié à la fois aux limites du support papier, à notre conception de la lisibilité, et à la concision à laquelle nous nous efforçons. En effet, la taille d'une feuille de papier limite le nombre de colonnes que l'on peut avoir avant que la transcription ne

devienne illisible. Nous jugeons que sur une feuille de papier de taille A4 orientée verticalement, trois colonnes représentent un bon compromis entre lisibilité et richesse multimodale. Concernant la concision, le fait d'avoir une seule colonne pour toutes les ressources visuelles nous oblige à penser à la notion de densité modale (Norris, 2004) et à ne capter que le ou les modes qui sont les plus denses. Si plusieurs modes interagissent ensemble, plusieurs lignes et/ou une description plus détaillée dans la colonne à droite s'imposent.

Enfin, nous reconnaissons qu'il est impossible de dire avec certitude que l'apprenant regarde l'image de l'enseignant ou qu'il lit les messages envoyés par clavardage.

5.4.3 Comptabiliser les annotations

Après avoir transcrit les phénomènes multimodaux sur ELAN, nous disposons de 28 fichiers de transcription, un fichier par séance. Un des principaux défauts d'ELAN est que ce logiciel ne nous donne pas une vision globale du corpus. Nos annotations étaient donc cloisonnées : il était impossible de voir des annotations simultanément sur l'ensemble des 295 séquences d'explication lexicale du corpus. Il fallait regarder une séance d'interaction à la fois. Or l'intérêt de notre corpus réside dans le fait que nous disposons d'un semestre entier d'interactions entre sept apprentis-enseignants et douze apprenants pour mener une étude de collection (Ten Have, 1999). Pour pallier ce problème, nous avons élaboré de nombreux tableaux sur Microsoft Excel afin d'avoir différents points de vue sur la totalité du corpus. Ces tableaux ont été alimentés par nos transcriptions sur ELAN. C'est à partir de nos tableaux que nous avons mené des analyses qui vont suivre dans les prochains chapitres.

Dans un premier temps, nous avons décrit librement et en détail tout ce qui se passe lors de chaque phase de chaque séquence d'explication lexicale. Ce premier tableau ressemble à la figure ci-dessous.

Semaine	Apprenti-enseignant	Item lexical	Compréhension/production	Déclenchée par	Ouverture enseignant	Ouverture apprenant	Noyau enseignant	Noyau apprenant	Clôture enseignant	Clôture apprenant
1-6	Adèle, Emilie, Etienne, Mélissa, Samia, Séverine, Victor	Quel est l'item lexical qui fait l'objet de la séquence ?	S'agit-il d'un problème de compréhension, de production ou des deux ?	Qui a initiée la séquence d'explication lexicale ?	Pour chacun des phases, nous décrivons librement et en détail tout ce qui se passe pour chacun des participants (phénomènes multimodaux et stratégies de communication et d'explication).					

Figure 7 : Premier tableau descriptif

Ce tableau contient 295 lignes correspondant aux 295 séquences d'explication lexicale relevées et son élaboration a nécessité plusieurs mois de travail. Il nous a permis de rassembler quelques informations préliminaires sur nos données et de comprendre le déroulement des séquences. Les descriptions qui en partie étaient codifiées par des termes récurrents nous ont permis d'effectuer des recherches par mot-clé. Par exemple, quand on recherche « visage d'incertitude » ou « geste » on retrouve toutes les séquences dont la description contient le mot recherché.

Après avoir réaffiné nos questions de recherche, nous avons élaboré un deuxième tableau. Les premières colonnes ont été conservées, mais les six dernières ont été remplacées par trois nouvelles colonnes. Au lieu de nous focaliser sur les phases, nous avons créé une colonne pour chaque question de recherche :

1. Comment les apprentis-enseignants se focalisent-ils sur la forme des items lexicaux ?
2. Comment les apprentis-enseignants se focalisent-ils sur le sens des items lexicaux ?
3. Comment les apprentis-enseignants impliquent-ils les apprenants dans les séquences d'explication ?

Notre deuxième tableau, contenant lui aussi 295 lignes, ressemble au tableau ci-dessous. Nous avons surligné en gris les colonnes redondantes avec le tableau précédent.

Semaine	Apprenti-enseignant	Item lexical	Compréhension/ production	Déclenchée par	Focalisation sur la forme	Focalisation sur le sens	Stratégies pour impliquer l'apprenant
1-6	Adèle, Emilie, Etienne, Mélissa, Samia, Séverine, Victor	Quel est l'item lexical qui fait l'objet de la séquence ?	S'agit-il d'un problème de compréhension, de production ou des deux ?	Qui a initiée la séquence d'explication lexicale ?	Pour ces trois grandes catégories, nous avons décrit librement et en détail la façon dont les apprentis-enseignants se sont focalisés sur la forme des items lexicaux, sur le sens, et comment ils ont impliqué les apprenants dans les séquences d'explication.		

Figure 8 : Premier tableau orienté vers les questions de recherche

Un changement épistémologique est à l'origine de la création de ce deuxième tableau. En effet, nous avons fait le choix de nous focaliser sur les trois questions de recherche ci-dessus car ce qui est intéressant d'un point de vue didactique n'est pas de savoir ce qui se passe lors de chaque phase d'une séquence d'explication, mais de savoir comment le savoir est transmis en visioconférence. Ce qui nous intéresse d'un point de vue didactique et d'un point de vue multimodal, c'est la façon dont l'apprenti-enseignant explique chaque facette de la connaissance d'un item lexical, et la façon dont l'apprenti-enseignant implique l'apprenant

dans la séquence d'explication. L'intérêt de notre recherche est l'étude de la transmission du savoir dans l'interaction et avec l'utilisation des ressources multimodales plutôt que l'étude de la structure globale des séquences les divisant en phases. Nous détaillerons da manière plus approfondie ces phénomènes dans les chapitres d'analyse.

Par la suite, nous avons élaboré trois tableaux spécialisés, un par chapitre d'analyse (hormis le chapitre 9 de comparaison). Chaque chapitre d'analyse traite d'une des trois questions de recherche précédemment distinguées (forme, sens, implication). Chacun de ces trois tableaux détaille l'usage des différents canaux de Visu pour remplir la fonction traitée par le chapitre d'analyse. La différence entre ces tableaux et les tableaux précédents est que les tableaux précédents répondent de manière très générale à nos questions de recherche. Les tableaux suivants sont plus spécialisés et nous permettent de savoir quels canaux sont sélectionnés et combinés par les apprentis-enseignants afin d'expliquer les items lexicaux et d'impliquer les apprenants dans les séquences d'explication. Nos trois tableaux spécialisés ont 295 lignes chacun et reprennent le modèle ci-dessous.

Semaine	Apprenti-enseignant	Item lexical	Compréhension/production	Déclenchée par	Microphone	Clavardage	Webcam	Complémentarité
1-6	Adèle, Emilie, Etienne, Mélissa, Samia, Séverine, Victor	Quel est l'item lexical qui fait l'objet de la séquence ?	S'agit-il d'un problème de compréhension, de production ou des deux ?	Qui a initiée la séquence d'explication lexicale ?	Comment les canaux de Visu sont-ils exploités pour se focaliser sur la forme de l'item lexical ? Sur le sens ? Pour impliquer l'apprenant dans la séquence d'explication ?			Quels canaux sont utilisés ensemble ?

Figure 9 : Trois tableaux spécialisés par chaque question de recherche

Enfin, après avoir élaboré nos trois tableaux spécialisés, nous avons créé trois derniers tableaux, un par question de recherche, afin de comptabiliser le nombre de séquences dans lesquelles chaque stratégie a été utilisée. Chacun de ces tableaux contient plusieurs onglets (feuilles), un par type de stratégie repérée. Par exemple, pour compter le nombre de séquences où les ressources visuelles de la webcam ont été utilisées pour se focaliser sur le sens d'un item lexical, nous avons créé une liste de toutes les séquences qui répondent à ce critère. Nous avons donc créé une liste pour chaque type de stratégie repéré pour chaque canal de Visu pour chacune de nos trois questions de recherche. La capture d'écran ci-dessous exemplifie notre façon de comptabiliser toutes les séquences d'explication lexicale où une ressource visuelle a été utilisée pour se focaliser sur le sens d'un item lexical.



	Geste	Photo		
2	Adèle S1 paradoxe	Adèle S5 camion-restaurant		
3	Adèle S1 travailler de chez soi	Mélissa S3 carte mentale		
4	Adèle S4 goûter	Mélissa S5 bobo		
5	Adèle S4 je m'en fous	Samia S5 quenelle		
6	Adèle S4 attractions	Samia S5 restauration mobile		
7	Adèle S4 équilibre	Victor S3 carte mentale		
8	Adèle S5 bobo 1	Victor S5 bobo		
9	Adèle S5 bobo 3			
10	Adèle S5 décontenancée			
11	Adèle S5 infusions			
12	Adèle S5 tartine			
13	Adèle S5 ferme			
14	Adèle S5 turnover/chiffre d'affaires			
15	Adèle S5 coûts			
16	Adèle S6 parcours			
17	Adèle S6 pour de vrai/fictif			
18	Emilie S1 caricature (Fiona)			
19	Emilie S1 caricature (Aiden)			
20	Emilie S2 arrondir les fins du mois			
21	Emilie S2 extrascolaire			
22	Emilie S4 pot			
23	Etienne S4 cache cache 1			
24	Mélissa S1 baïller			
25	Mélissa S2 collègue			
26	Mélissa S2 chaud			
27	Mélissa S3 signer ma convention			

Figure 10 : Listes de séquences d'explication par stratégie repérée

Grâce à ces tableaux, nous pouvons mesurer l'usage de chaque canal de Visu pour chaque stratégie d'explication que nous avons repérée. Un autre avantage de ces listes est que la sélection d'exemples est facilitée. Au lieu d'effectuer des recherches par mot-clé, il suffit de consulter la liste et d'en sélectionner un exemple. Tandis que les tableaux précédents nous permettent de voir l'interaction entre les canaux de Visu, de voir par exemple qu'un geste se combine à une définition orale, ces derniers tableaux se focalisent sur un canal à la fois.

Pour résumer, nous avons créé des tableaux pour pallier le défaut principal d'ELAN : son incapacité à donner une vision globale du corpus. Huit tableaux ont été créés en tout : un tableau préliminaire et exhaustif, un tableau orienté par nos questions de recherche, puis six tableaux plus spécialisés, deux par question de recherche.

5.5 Conclusion

Nous avons décrit notre positionnement par rapport à nos données, notre façon de repérer les séquences d'explication lexicale, et notre façon d'annoter les phénomènes multimodaux. Nous avons repéré 295 séquences d'explication lexicale qui se repartissent presque également entre des problèmes de compréhension et des problèmes de production. Peu de séquences relèvent à la fois d'un problème de compréhension et d'un problème de production.

Après avoir repéré les séquences d'explication lexicale, nous avons annoté les phénomènes multimodaux en utilisant ELAN. Nous avons annoté les postures, les regards, les

gestes, et les expressions faciales. Nous avons justifié ces choix méthodologiques. Afin d'avoir des points de vue globaux sur le corpus, nous avons élaboré plusieurs tableaux à partir desquels nous avons initié nos analyses. Nos données ont donc été traitées à plusieurs niveaux et par plusieurs logiciels. Le tableau ci-dessous complète la Figure 10 du chapitre précédent en page 134.

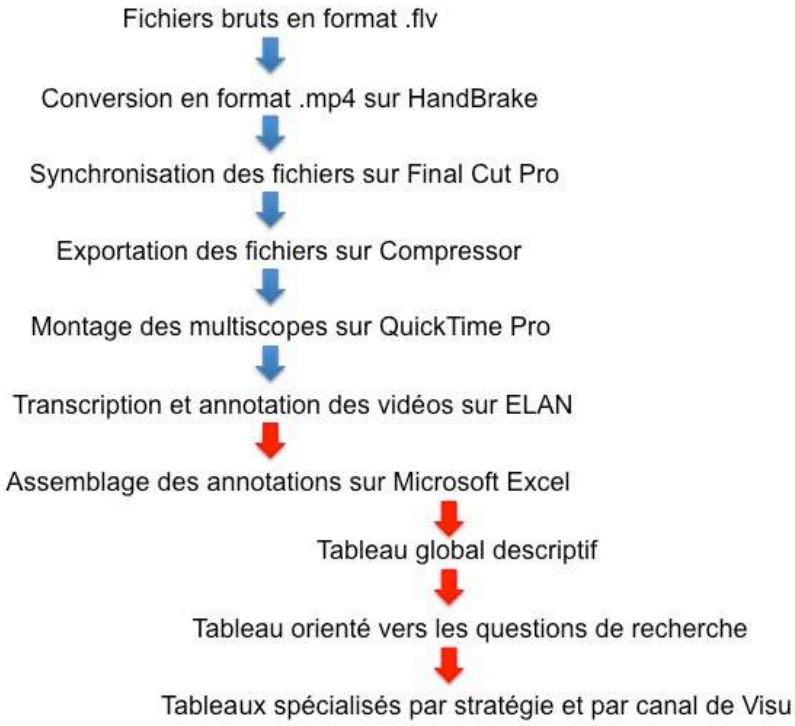


Figure 11 : Transformation et annotation des données

Comme le tableau ci-dessus le met en évidence, nos données ont été profondément transformées. Ces transformations ont des répercussions sur les analyses. Le choix de fixer l'image de l'enseignant deux fois plus grande que celle de l'apprenant a comme conséquence qu'en regardant la vidéo sur ELAN, les mouvements des enseignants sont plus perceptibles que ceux des apprenants. Ce choix reflète la priorité que nous donnons à l'analyse des pratiques multimodales des enseignants. Pour l'annotation de nos données, le fait d'utiliser ELAN a pour conséquence que l'on visualise la multimodalité comme étant une juxtaposition de plusieurs modes prédéfinis. Or, nous avons expliqué dans le chapitre 3 que selon le principe d'intégration des ressources (Baldry & Thibault, 2006), les modes interagissent les uns avec les autres pour créer un effet qui est plus complexe que leur simple addition. Nous nous appuyons sur nos connaissances en tant qu'utilisateur de plateformes de visioconférence et en tant que membre de la communauté des enseignants pour interpréter cette complexité. Puis, le fait de représenter nos annotations sur des tableaux à trois colonnes a comme conséquence que l'on ne se focalise que sur une image à la fois, alors que l'interaction est une

vidéo avec 3 images par seconde. Notre choix d'élaborer des tableaux descriptifs et quantitatifs divise les interactions complexes en petites unités d'analyse qui ont l'avantage de nous aider à répondre à nos questions de recherche, mais qui risquent de ne pas prendre en compte des éléments contextuels plus larges (Rivens Mompean & Cappellini, 2015).

Aucune analyse n'est vraiment objective car chacune de ces transformations a été le résultat d'un choix méthodologique. Notre recherche est fortement orientée par ces choix méthodologiques mais aussi par nos compétences techniques. Avant de commencer cette thèse, nous avions de l'expérience de tous les outils numériques utilisés sauf Final Cut Pro. Si nous avions transcrit nos données en utilisant un autre logiciel qui donne une vue globale du corpus, comme Transana (Lascar et Valéro, 2016), nous n'aurions pas créé nos tableaux de la même façon. Une certaine maîtrise des outils technologiques est nécessaire pour mener des études dans le domaine des humanités numériques. L'observateur ne peut donc être qu'une personne qui a une certaine maîtrise de la technologie.

Nous allons exploiter nos annotations dans les prochains chapitres d'analyse afin de savoir comment les apprentis-enseignants expliquent les items lexicaux et impliquent leurs apprenants dans les séquences d'explication lexicale.

Partie III : Analyses des données

Dans cette partie, nous nous appuierons sur le corpus ISMAEL dont la construction a été détaillée dans le chapitre 4 et dont l'annotation multimodale a été expliquée dans le chapitre 5 afin de répondre à nos questions de recherche. Cette partie se divise en cinq chapitres.

Le chapitre 6 répondra à la question suivante : comment les apprentis-enseignants se focalisent-ils sur la forme des items lexicaux ? Ce chapitre nous permettra de faire état des moyens multimodaux par lesquels les apprentis-enseignants dirigent l'attention des apprenants vers la forme des items lexicaux. L'importance de la focalisation sur la forme a été mentionnée dans le chapitre 2 et sera revisitée au début du chapitre avant de passer aux analyses.

Le chapitre 7 répondra à la question suivante : comment les apprentis-enseignants se focalisent-ils sur le sens des items lexicaux ? Afin de faciliter la présentation des analyses, ce chapitre sera divisé en deux grandes sections : les explications voco-verbales et les explications ostensives.

Le chapitre 8 répondra à la question suivante : comment les apprentis-enseignants impliquent-ils les apprenants dans les séquences d'explication lexicale ? Compte tenu de l'importance de la négociation du sens, de l'implication personnelle, et de la participation active des apprenants, (cf. chapitre 2), il reste à savoir par quels moyens multimodaux les apprentis-enseignants remplissent leur obligation didactique d'assurer une implication cognitive et socioaffective des apprenants.

Le chapitre 9 donnera lieu à une occasion rare de comparer cinq apprentis-enseignants dans des conditions similaires expliquant le même item lexical à leurs apprenants. Dans une perspective de comparaison interindividuelle, nous comparerons la conduite de plusieurs séquences d'explication lexicale à partir de critères définis au début du chapitre. Les analyses s'articuleront autour de neuf explications de deux items lexicaux portant une charge culturelle importante : « bobo » et « goûter d'anniversaire ».

Enfin, le chapitre 10 sera l'occasion de synthétiser nos résultats, de formuler des propositions didactiques et de discuter des limites de notre étude.

Une sélection d'extraits vidéo et d'annotations multimodales se trouve sur notre site compagnon :

http://www.icar.cnrs.fr/membre/bholt/site-compagnon-these/

Chapitre 6 : La focalisation sur la forme des items lexicaux

6.1 Introduction

Un item lexical est constitué de plusieurs facettes qu'il faut maîtriser avant de prétendre « connaître » l'item entièrement. Les trois grandes catégories sont la forme, le sens et l'usage (Nation, 2013). Ce chapitre se donne pour objectif d'analyser la façon dont les apprentis-enseignants guident la focalisation sur cette première grande facette : la forme. On sait déjà que la focalisation sur la forme d'un item lexical est essentielle en particulier lors de l'ouverture d'une séquence d'explication lexicale (Fasel Lauzon, 2014) quand l'objet de la séquence est repéré, décontextualisé et problématisé. On connaît aussi plusieurs techniques linguistiques et multimodales par lesquelles un enseignant peut décontextualiser un item lexical et mettre en exergue sa forme. Il reste à savoir si l'interaction pédagogique en visioconférence rend visible d'autres fonctions de la focalisation sur la forme d'un item lexical, et s'il y a de nouvelles techniques multimodales pour le faire. Pour répondre à ces questions, nous allons dans un premier temps faire un survol de la littérature sur l'intérêt pédagogique de la focalisation sur la forme des items lexicaux et sur les techniques de décontextualisation. Dans un second temps, nous allons analyser la focalisation sur la forme lors des problèmes de compréhension et de production.

6.2 Focalisation sur la forme : un rapide état de l'art

Dans cette première section, nous allons rappeler l'importance pédagogique de la focalisation sur la forme des items lexicaux avant de faire un survol de la littérature afin de dégager des techniques connues de cette focalisation.

6.2.1 Utilité pédagogique de la focalisation sur la forme

Les avantages pédagogiques associés à la focalisation sur la forme sont bien connus. La focalisation sur la forme permet notamment aux processus d'acquisition langagière d'être plus rapides et plus exactes, à condition que la focalisation sur la forme soit implémentée correctement, c'est-à-dire sans qu'elle ne nuise pas à la communication et à la tâche (Long,

1991). Ellis, Basturkmen, & Loewen (2001) listent cinq critères de la focalisation sur la forme :

1. Elle arrive dans un discours dont le premier objectif est communicatif ;
2. Elle est observable, c'est-à-dire interactionnelle ;
3. Elle est incidente, c'est-à-dire non planifiée ;
4. Elle est transitionnelle, c'est-à-dire qu'elle a une durée limitée avant de reprendre l'objectif communicationnel ;
5. Elle est large, dans le sens où plusieurs types d'éléments peuvent faire l'objet de focalisation (items lexicaux, points grammaticaux, etc.)

Nous retenons tous ces critères sauf le troisième et le cinquième. En ce qui concerne le troisième point, certains épisodes de focalisation sur la forme peuvent être planifiés par les apprentis-enseignants par exemple quand il s'agit de vérifier la compréhension d'un mot-clé. En ce qui concerne le cinquième point, bien que plusieurs types d'éléments puissent faire l'objet de focalisation, nous ne nous focaliserons dans ce chapitre que sur les items lexicaux qui font l'objet d'une focalisation manifeste. Certes, comme le rappellent ces auteurs, il peut y avoir focalisation sur la forme si l'apprenant se focalise mentalement sur un item lexical sans aucun indice observable. De même, s'il y a focalisation manifestée sur la forme, cela ne veut pas dire forcément que l'apprenant se focalise mentalement sur la forme.

Afin d'apprendre, le repérage de formes est absolument essentiel, car l'apprentissage ne peut pas avoir lieu de manière inconsciente (Schmidt, 1990). Selon cet auteur, ceci est vrai pour tous les aspects de l'apprentissage, y compris pour l'apprentissage du vocabulaire. L'apprentissage de formes qui ne sont pas centrales pour la tâche en cours est possible pour les enfants, mais non pas pour les adultes. Les apprenants doivent clairement comprendre qu'il existe une focalisation sur un item langagier car on apprend ce sur quoi on se focalise (Barcroft, 2006 ; Hill & Laufer, 2003). Le repérage est un processus cognitif qui se manifeste par le fait de porter l'attention sur un item lexical et de se rendre compte de son utilité, sa pertinence à la tâche en cours et/ou le fait qu'il soit inconnu. C'est l'un des éléments qui caractérisent l'ouverture d'une séquence d'explication lexicale : c'est le moment où l'item lexical qui va par la suite faire l'objet d'une explication est repéré, décontextualisé et problématisé (De Gaulmyn, 1991; Ducancel, 1991; Fasel Lauzon, 2014; Gülich, 1990). Une séquence d'explication lexicale ne peut donc pas avoir lieu sans focalisation sur la forme. Enfin, la mise en exergue de la forme d'un item lexical s'impose aussi car la capacité à

apprendre des mots dépend fortement de la représentation phonologique mentale que se font les apprenants (Nation, 2013).

Compte tenu de l'importance de la perception et du repérage de la forme de l'item lexical, l'enseignant se doit « d'augmenter au maximum le degré de saillance, de 'manifestness', des unités lexicales considérées comme termes-clés pour en permettre l'appréhension à l'alloglotte » (Lüdi, 1991 : 213). Cette mise en exergue d'items lexicaux a déjà été étudiée dans des cours en présentiel. Flowerdew (1992), qui a étudié 329 définitions données dans des cours magistraux de biologie et de chimie à des locuteurs non-natifs, parle d'accentuations linguistiques et paralinguistiques[1] pour introduire un terme faisant l'objet d'une explication.

En ce qui concerne la mise en exergue d'un item lexical, Nation (2013) explique que la décontextualisation peut se faire par plusieurs moyens :

1. L'apprenant constate, lors d'une tâche réceptive (écoute/lecture), la présence d'un item lexical qu'il ne connaît pas ou dont l'emploi est nouveau ou différent.
2. L'enseignant souligne un item lexical ou le met en exergue. Nation donne l'exemple d'un enseignant dans une salle de classe qui écrit un item lexical au tableau.
3. Le sens d'un item lexical est négocié entre apprenants ou entre apprenant(s) et enseignant.
4. L'enseignant explique un item lexical en donnant une définition, un synonyme ou une traduction.

Ce chapitre se propose de traiter ces deux premiers types de décontextualisation. Le quatrième point sera abordé dans le chapitre suivant, qui portera sur la façon dont le sens d'un item lexical est expliqué, et la façon dont l'enseignant renforce le lien entre la forme et le sens. Le troisième point sera abordé dans le chapitre 8 qui analysera la façon dont l'enseignant implique les apprenants dans une séquence d'explication lexicale.

6.2.2 Décontextualiser un item lexical

Quand un item lexical est décontextualisé, il obtient souvent le statut autonymique : « Prenez un signe, parlez-en, et vous aurez un autonyme » (Rey-Debove, 1997: 144). Le passage d'un mot ordinaire en mot autonyme peut arriver à tout moment. Quand on parle d'un

[1] « linguistic and paralinguistic emphasizers » (Flowerdew, 1992 : 290).

mot, on monte dans la hiérarchie des langages : on parle de la langue[2]. Un autonyme est « un objet dont la matière est linguistique (les lettres, les phonèmes). Dans la conscience de l'encodeur, c'est un objet connu pour être un signe » (Rey-Debove, 1997 : 147). Cette auteure (1997 : 33-34) en donne cette définition :

> « Le mot autonyme signifie le signe de même signifiant dont il est le nom, et ne signifie que cela ; le mot métalinguistique signifie tout ce qui est langagier, excepté le signe de même signifiant dont il serait le nom ; /*table*/ signifie « mot *table* (avec sa propre expression et son propre contenu) » mais *morphème* signifie « unité signifiante minimale (dont l'expression et le contenu ne sont pas déterminés) ».[3]

Ainsi, si un enseignant dit « /table/ est un mot masculin », /table/ » est autonymique car il signifie *mot table* et non pas l'objet dont il est le nom. Cette autonymisation s'impose dès lors que l'on veut parler d'un objet langagier comme élément d'étude, par exemple parler d'un item lexical lors d'une séquence d'explication. Il existe plusieurs façons de marquer le statut autonymique d'un objet langagier.

Pour le linguiste, une manière de marquer le statut autonymique d'un mot écrit ou d'une phrase est de l'entourer de cotices (barres obliques), comme le fait Rey-Debove. Par exemple, /table/ signifie « mot table ». Rey-Debove (1997 : 29) précise que les mots et leurs autonymes sont homophones mais ne sont pas le même mot en expliquant :

> « (1) que les mots autonymes sont homonymes des mots qu'ils désignent ; *table* et /*table*/ sont des mots différents, (2) que les mots autonymes servant à parler des mots d'un lexique l_1 d'une langue L_1, sont tous des noms, et appartiennent à ce lexique dans la mesure où la compétence pour *table* entraîne obligatoirement la compétence pour /*table*/. » (29).

Les mots autonymes ont certaines caractéristiques qui facilitent leur repérage. Quand un mot devient autonyme, il est obligatoirement nominalisé, quel que soit son statut :

> « Le signe autonyme, quel que soit son signifiant, est un nom (noun). Il ne saurait en être autrement, car parler d'un signe c'est le prendre comme sujet du discours, et le sujet du discours possède, dans toutes les langues qui ont des noms, la fonction nominale. Pour parler d'un verbe, d'un adverbe, il faut les nominaliser » (Rey-Debove, 1997 : 64).

Soit une séquence d'explication lexicale dont l'objet est le verbe *aller*. Le verbe devient nominalisé : /*Aller*/ est un nom qui signifie *verbe aller*.

Quand un mot devient autonyme, il devient un signe dont le signifié est un signe. Autrement dit, il devient signe d'un signe. Rey-Debove (1997 : 94) en donne un exemple : « /*Libraire est dissyllabique*/ » où *libraire* est un autonyme qui possède à la fois un signifiant et un signifié. « /*Libraire*/ signifie « Nom dont le signifiant graphique est l-i-b-r-a-i-r-e, le

[2] Pour parler d'une lange L_n, on utilise la langue L_{n+1}. Pour parler de *la langue* L_{n+1}, on utilise *la langue* L_{n+2}, et ainsi de suite.

[3] Dans le discours métalinguistique de cette auteure, un mot entouré de cotices est autonymique. Ainsi, dans cette citation, /table/ est autonymique tandis que *morphème* ne l'est pas. Par exemple, /aller/ est un autonyme qui signifie « le verbe aller ».

signifiant phonique [libRɛR] et qui signifie 'Personne qui vend des livres' » (95). Donc le signe qui est signifié par l'autonyme comprend à la fois un signifiant et un signifié. L'autonyme a cette double fonction de « transforme[r] en nom toute partie du discours », et de « transforme[r] en signe tout ce qui n'est pas signe » (Rey-Debove, 1997 : 101). Les autonymes sont à la fois des noms et des signes.

Il faut préciser qu'une phrase entière peut aussi être autonymique, par exemple une phrase-exemple d'un linguiste ou d'un enseignant. L'autonyme composé de plusieurs mots ou morphèmes est « un morphème original unique » (Rey-Debove, 1997 : 149) quel que soit le nombre de mots qui le composent. Par exemple, l'autonyme « /*Je pars demain*/ » est un autonyme au même titre que « /*libraire*/ » (149). Le sens de l'autonyme ne dépend pas du sens des mots qui le composent. Ainsi, une expression idiomatique ou une collocation peut devenir autonymique lors d'une séquence d'explication lexicale.

Le passage d'un mot ordinaire en mot autonyme (ou phrase ordinaire en phrase autonyme) a lieu souvent lors de d'une séquence d'explication lexicale. Ce passage sert à la fois à : 1) marquer l'item lexical qui fait l'objet de la séquence 2) l'extraire de son contexte (même si on admet qu'aucun signe ne peut être totalement décontextualisé)[4], et 3) passer d'un discours ordinaire à un discours sur la langue.

Avant de procéder à des analyses didactiques sur la focalisation sur la forme, nous allons faire le point sur différents moyens par lesquels un apprenti-enseignant peut mettre en exergue la forme d'un item lexical lors d'une séquence d'explication lexicale, qui est une forme de discours sur la langue. Ces stratégies seront illustrées à l'aide d'exemples.

6.2.2.1 Les marques voco-verbales

Il existe plusieurs moyens de décontextualiser, d'autonymiser un item lexical en utilisant sa voix. Ces moyens incluent la prononciation plus lente, un ton de la voix élevé, l'emphase, la saillance vocalique, l'allongement de syllabes (Chaudron, 1982). Pour cette raison, les exemples suivants ne seront pas transcrits de manière multicanale car seulement le microphone est utilisé.

6.2.2.1.1 Les logonymes

La *didacticité* peut être abordée, selon Cicurel (2011 : 21), de trois manières différentes, y compris « selon une définition *formelle* s'attachant à des marques linguistiques

[4] « L'idée de signe isolé, sans contexte, etc., est un leurre. Tout signe est, de façon sensible ou mentale, quelque part ; ou dans un contexte mondain : contexte langagier, contexte sémiotique non langagier, contexte situationnel (état des choses) ou dans un contexte métalinguistique présentant les mêmes caractères, contexte métalangagier, métasémiotique, situationnel (état du discours) » (Rey-Debove, 1997 : 90).

montrant qu'il y a transmission de savoir (reformulations, définitions, mots métalinguistiques…) ». Nous allons nous focaliser sur les mots métalinguistiques car c'est une façon très efficace et économe de décontextualiser un objet langagier.

Rey-Debove (1997) rappelle que le langage possède une « fonction métalinguistique », c'est-à-dire que le langage peut être utilisé pour parler du langage. L'auteure soutient que d'un point de vue logique, le métalangage doit être plus complexe que le langage qu'il décrit. Une langue L_n est donc incluse dans une langue L_{n+1} qui la décrit. Les enseignants de langue ont besoin de cette compétence métalinguistique pour leur permettre d'enseigner la matière qu'ils enseignent : « La compétence linguistique permet de produire des phrases acceptables sur le monde, la compétence métalinguistique, de produire des phrases acceptables sur la langue, notamment celles qui affirment que les phrases sur le monde sont ou non acceptables » (Rey-Debove, 1997 : 21). Les enseignants passent souvent de la langue L_n à la langue L_{n+1} pour expliquer des phénomènes langagiers, et ce passage marque souvent le début d'une séquence d'explication lexicale. Les enseignants ont donc intérêt à prendre conscience de ce passage et à le marquer afin que les apprenants le repèrent. Pour parler du langage, on peut employer des mots linguistiques. Un mot linguistique est « un mot qui, à quelque degré, enferme la notion de langage dans son signifié, sans être un mot autonyme » (Rey-Debove, 1997 : 32). Une façon de le faire est d'utiliser des logonymes.

Les logonymes sont utilisés par les apprentis-enseignants pour qualifier un item lexical d'objet langagier dont ils veulent vérifier la compréhension ou qu'ils souhaitent décontextualiser avant de passer au noyau de l'explication. Marra & Pallotti (2006 : 128) disent que :

> « Le lexique logonymique contient les formes à travers lesquelles la langue elle-même, dans son utilisation quotidienne, se définit et « auto-certifie » sa dimension linguistique et communicative (Silvestri, 2000). Appartiennent à la catégorie des logonymes tous les mots qui désignent des states, des modalités, des produits et des actions de la communication humaine. Par conséquent, en font partie tous les verbes à travers lesquels nous exprimons la production et la réception linguistiques (comme 'dire', 'parler', 'chuchoter', 'écrire', 'écouter', 'lire') et tous les produits de cette activité (comme 'mot', 'chuchotement', 'écriture', 'texte', 'écoute', 'lecture' mais également 'langue' avec ses différentes manifestations : 'anglais', 'français', etc.). Enfin, on qualifiera également de logonymiques les mots qui désignent les producteurs et récepteurs de cette activité : 'locuteur', 'orateur', 'auditeur', etc. »

L'usage de logonymes signale la focalisation sur un item langagier et introduit souvent une explication lexicale. Il signale aussi une focalisation sur les normes langagières : Cicurel (2011 : 70) explique que « on dit » « est un verbe introducteur du dire qui renvoie à un énonciateur représentant la norme langagière ». L'usage de logonymes est tellement efficace comme marqueur et agent d'ouverture de séquences d'explication que Fasel Lauzon (2014) a

recensé les logonymes comme « expliquer » et « explication » pour repérer et identifier les 241 séquences d'explication qu'elle a analysées pour son étude. Dans l'exemple suivant, Séverine utilise le logonyme « mot » pour décontextualiser l'item lexical dont elle vérifie la compréhension :

Transcription de la parole		
1	SEV	la ((rire)) la retraite/ (.) tu comprends/ (.) le mot retraite/ ou pas\

Exemple 1 : Séverine utilise le logonyme « mot »

Dans cet exemple, le logonyme « mot » permet à Séverine de guider l'attention de son apprenante vers l'item lexical « retraite ».

Pour compter l'usage de logonymes dans notre corpus, nous avons créé un tableau dans lequel nous avons marqué la focalisation sur la forme pour chaque séquence d'explication lexicale. Ensuite, nous avons compté le nombre de fois que chaque logonyme apparaît dans le corpus. Nous avons repéré les logonymes suivants : *parler*, *mot*, *dire*, *français*, *appeler*, *expliquer*, *verbe*, *anglais*, *expression*, *formulation*, *entendre*, *principe*, *décrire*, *terme*, *nom*, et *espagnol*. Les logonymes les plus utilisés sont « dire » et « mot ». L'utilisation de logonymes se résume comme suit :

	Adèle	Mélissa	Victor	Samia	Étienne	Séverine	Emilie	Total
Parler	1		3	3			2	9
Mot	7	11	10		1	3	6	38
Dire	15	6	4	4	2	9	6	46
Français	3	2	8	1		3		17
Appeler	4	2	3	3		1		13
Expliquer	1		1				1	3
Verbe	1	2						3
Anglais	1		4			1		6
Expression	1		5			1	3	9
Formulation							1	1
Entendre							1	1
Principe					1			1
Décrire		1						1
Terme		2						2
Nom		1						1
Espagnol		1						1
Total	34	28	38	11	4	18	19	152

Figure 1 : Utilisation de logonymes par les apprentis-enseignants

Les comparaisons directes entre les apprentis-enseignants sont à éviter parce que le nombre de séances qui ont été transcrites ne sont pas les mêmes. Pour Étienne et Séverine, ces chiffres représentent l'utilisation de logonymes pour une seule séance. Pour Emilie, c'est quatre séances. Pour Victor et Samia, c'est cinq. Pour Adèle et Mélissa, ces chiffres représentent six séances. La colonne la plus intéressante scientifiquement est donc celle de droite qui montre la totalité.

6.2.2.1.2 Les dislocations

L'analyse conversationnelle s'intéresse à ce dont les interlocuteurs parlent, et à ce sur quoi ils se focalisent. La recherche issue de ce domaine a identifié plusieurs moyens linguistiques par lesquels les interlocuteurs s'orientent mutuellement vers un topic. Ces réorientations d'attention nous intéressent en tant que didacticien dès lors que le topic s'agit d'un phénomène didactique observable comme les séquences d'explication lexicale. Un des moyens linguistiques par lesquels les interactants peuvent marquer un changement de focus ou de topic est la dislocation. Une dislocation permet de séparer un élément du reste de la phrase. Les dislocations en français sont très fréquentes : « Le français, ça se disloque » (Horlacher, 2015 : 11). La dislocation, qui est « un procédé d'emphase, relevant de la syntaxe expressive » est expliquée ainsi par Horlacher (2015 : 12) :

> « La dislocation est liée à l'organisation de l'information et permet de séparer le thème du reste de la phrase. Selon cette interprétation, la dislocation – dans sa version gauche – permet de satisfaire un principe pragmatique de base en français selon lequel un référent doit être posé dans le centre de l'attention des interlocuteurs avant qu'il n'en soit prédiqué quelque information à son sujet. Dans sa version droite, le principe évoqué ici permet de prédiquer quelque chose au sujet d'un référent, qui ne sera thématisé que dans un second temps. La dislocation à droite obéit donc aussi à ce principe de « dissocier en deux phases énonciatives, ce qui, sinon, serait effectué en un seul mouvement » (Apothéloz 1997 : 195-196).

En didactique des langues, cet instrument linguistique d'emphase peut être employé à bon escient afin d'introduire un item lexical qui va par la suite faire l'objet d'une explication ou de mettre en exergue un item lexical qui vient d'être expliqué.

6.2.2.1.2.1 La dislocation à gauche

Pekarek Doehler (2001 : 180) définit la dislocation à gauche comme « une construction syntaxique dans laquelle un syntagme nominal (SN) extrait de la clause apparait à gauche, suivi d'une clause comprenant un pronom clitique (dit 'élément de reprise') qui coréfère avec le SN […] ». Cette auteure explique que :

> « Selon la littérature à ce sujet, cette construction syntaxique a pour fonction fondamentale d'introduire ou de réintroduire un référent dans le discours qui n'est pas au centre de l'attention du destinataire, mais qui est néanmoins accessible pour ce dernier […] ; elle sert à promouvoir un référent au statut de topic (Lambrecht, 1987). »

La dislocation à gauche a des fonctions cognitives et interactives qui la rendent très utile pour introduire un item lexical. « La fonction cognitive de la dislocation à gauche consiste à signaler au destinataire que les énoncés suivant la dislocation sont à interpréter comme éléments d'un nouveau segment de discours » (Pekarek Doehler, 2001 : 181). Cette auteure (2001 : 190) explique que « Sur le plan interactif, elle accomplit une fonction cognitive de cadrage de ce qui suit, signalant notamment que cette suite effectue un départ (mais non pas une rupture) par rapport à ce qui précède ». La dislocation à gauche « est un instrument dont se servent les interlocuteurs pour coordonner leur orientation réciproque vers un segment thématique et/ou interactif nouveau, mais rattaché au précédent » et un instrument qui peut être perçu « comme la grammaticalisation d'une fonction pragmatique, à savoir celle de promouvoir un élément au statut de topic » Pekarek Doehler (2001 : 190). Souvent, ce nouveau segment thématique est le noyau d'une séquence d'explication lexicale, le topic étant l'item lexical.

Dans l'exemple suivant, Séverine emploie une dislocation à gauche pour introduire l'item lexical avant son explication :

Transcription de la parole	
1 SEV	alors hygiène c'est (0,2) hygiène c'est la propreté

Exemple 2 : Dislocation à gauche pour introduire un item lexical

La dislocation à gauche permet d'attirer l'attention vers l'item lexical « hygiène » qui est défini immédiatement après.

6.2.2.1.2.2 La dislocation à droite

Tandis que le changement de topic et la réintroduction d'un topic sont normalement attribués à la dislocation à gauche, la dislocation à droite est souvent utilisée pour maintenir un topic. Contrairement à la dislocation à gauche, qui peut être utilisée pour réduire la charge mémorielle d'un interlocuteur, la dislocation à droite est plus coûteuse en charge mémorielle parce que qu'elle crée une séquence en suspens. Horlacher (2015 : 45) donne l'exemple suivant de ce type de « *nominativus pendens* » : « Je ne suis pas d'accord, les vacances n'importe quand ». La dislocation à droite arrive après-coup pour clarifier, désambiguïser ou ajouter des informations manquantes. C'est une stratégie utile lors des interactions didactiques : « Chez l'apprenant d'une L2, notamment, la surspécification des entités et le marquage répété du topic par des structures disloquées constituent des moyens pour déjouer les risques d'ambiguïté et favoriser l'intercompréhension » (Pekarek Doehler, 2004 : 9).[5] Bien

[5] Cette page correspond à la pagination du manuscrit auteur.

que la fonction la plus citée de la dislocation à droite soit celle de la maintenance d'un topic, Horlacher (2015) s'inscrit en faux contre l'idée selon laquelle une dislocation à droite ne puisse pas introduire un nouveau topic. Elle peut être utilisée pour introduire un nouveau topic ou pour enchaîner sur un topic précédent, ou pour en recycler un. Nous considérons donc les deux types de dislocations comme des moyens valables pour gérer les séquences d'explication lexicale.

Dans l'exemple suivant, Samia utilise une dislocation à droite pour réintroduire un item lexical afin d'en vérifier la compréhension.

	Transcription de la parole
1 SAM	et faire des propositions/ comme par exemple euh je sais pas moi euh (0,4) une formule de repas (1,2) vous savez ce que c'est les formules de repas/

Exemple 3 : Dislocation à droite pour vérifier la compréhension d'un item lexical déjà introduit

L'utilisation de dislocations se fait comme suit :

	Adèle	Mélissa	Victor	Samia	Étienne	Séverine	Emilie	Total
A Gauche	5	6	5	5		6	4	31
A Droite	3	1	9	11		2		26
Total	8	7	14	16	0	8	4	57

Figure 2 : Utilisation de dislocations

On constate une fréquence similaire de dislocations à gauche et à droite, ce qui corrobore l'avis de Horlacher que les deux types de dislocations peuvent introduire et maintenir un topic.

6.2.2.1.3 La lemmatisation

La lemmatisation est une façon de standardiser les diverses formes d'un item lexical que l'on peut rencontrer dans les discours. Mortureux (2008 : 14) explique que « L'opération qui consiste à ramener à une forme unique les formes fléchies des mots variables (noms, adjectifs et verbes, tels qu'ils figurent dans les discours) s'appelle la *lemmatisation* ». C'est une façon d'universaliser un item lexical, car une des caractéristiques d'une séquence d'explication lexicale est de permettre à l'apprenant d'acquérir des connaissances générales à partir d'une instanciation précise. Dans l'exemple suivant, suite à une définition, Victor renforce la perception de la forme orale de l'item lexical en donnant la forme lemmatisée de la collocation « faire la grève ».

Transcription de la parole		
1	VIC	c'est quand les gens/ (.) arrêtent de de travailler/ (.) et ils sortent dans la rue/ pour scander des (.) des slogans:/ et\ (1.0) (et) vouloir de de meilleurs conditions de travail\ (0,6) faire la grève\

Exemple 4 : Usage de la lemmatisation suite à une définition

Nous avons trouvé deux exemples de lemmatisation dans le corpus, un par Victor et l'autre par Mélissa.

6.2.2.1.4 Les pauses et saillance vocalique

On peut introduire l'autonymie par des marques phoniques comme des pauses, emphases, mélodies, et intonations. L'isolement phonique est caractérisé par une pause ou un manque de liaison avec ce qui précède et ce qui suit (Rey-Debove, 1997). Un apprenti-enseignant peut diriger l'attention de l'apprenant vers un item lexical par un de ces moyens. Dans l'exemple suivant, Adèle donne la phrase autonymique et lemmatisée « travailler de chez moi » à ses apprenantes. Cela faisant, elle combine les logonymes « dire » et « français » avec des pauses qui entourent chaque composant de la collocation.

Transcription de la parole		
1	ADL	quand quand tu veux dire il travaille euh à distance/ si vous voulez dire ça en français/ (0,6) tu dis/ (.) euh (.) je travaille/ (0,4) de/ (0,3) chez moi\

Exemple 5 : Usage de pauses pour mettre en évidence une collocation

En entourant les composants (mots linguistiques) de la collocation de pauses, Adèle se focalise sur la collocation « travailler de chez moi » et améliore ainsi les chances que les apprenantes s'en fassent une représentation phonologique forte.

Dans l'exemple suivant, la forme orale du mot « clown » en français fait défaut à l'apprenant. Quand Étienne comprend ce que son apprenant cherche à dire, il donne la forme orale de l'item lexical et met l'emphase sur celle-ci par saillance vocalique.

Transcription de la parole		
1	ETN	ah un CLOWN\

Exemple 6 : Étienne lève la voix

Au total, 27 usages de saillance vocalique pour marquer un item lexical ont été observés.

Adèle	Mélissa	Victor	Samia	Étienne	Séverine	Emilie	Total
8	6	6	3	2	1	1	27

Figure 3 : Saillance vocalique pour se focaliser sur la forme orale d'un item lexical

Une autre façon d'introduire un autonyme est d'abandonner la règle de liaison, par exemple « l' » en français : « La tendance, dans ce cas, est de traiter la séquence autonyme commençant par une voyelle (phonique), comme si elle commençait par une consonne : /Le *ou* a été remplacé par *et*/ (et non pas, **l'ou*). » (Rey-Debove, 1997 : 65). Lorsqu'un autonyme est introduit avec la liaison, le risque d'ambiguïté augmente. La liaison est abandonnée plus facilement lorsque l'autonyme n'est pas un nom, par exemple « /ce *on* est trop anonyme/, /la prononciation de *asseoir*/ » (Rey-Debove, 1997 : 66). L'auteure plaide pour que l'on abandonne la liaison pour rendre plus visible l'autonyme. Nous n'avons pas trouvé un exemple dans notre corpus.

6.2.2.2 Marques multimodales

Les marques multimodales impliquent plusieurs canaux. Celles-ci nous intéressent en visioconférence car les utilisateurs disposent de quatre canaux (la webcam, le microphone, le clavardage et l'envoi de documents) qui véhiculent plusieurs ressources sémiotiques chacun (cf. chapitre 3). L'orchestration multimodale par les apprentis-enseignants est d'un grand intérêt en apprentissage de langues médié par ordinateur (Guichon, 2013).

6.2.2.2.1 Renforcement par clavardage

Lors des interactions pédagogiques médiées par ordinateur, le clavardage est un outil de grande valeur car il permet à l'enseignant de redoubler par écrit des messages oraux et d'intervenir sans interrompre l'apprenant (Dejean-Thircuir, Guichon, & Nicolaev, 2010 ; Develotte, Guichon, & Kern, 2008 ; Develotte & Mangenot, 2007 ; Guichon, Bétrancourt, & Prié, 2012 ; Lamy, 2006 ; Nicolaev, 2010 ; Yanguas, 2010). C'est un bon moyen de trouver un équilibre entre communication et focalisation sur la forme dont parle Long (1991). Cet outil est souvent utilisé par les apprentis-enseignants lors des séquences d'explication lexicale pour mettre en exergue la forme écrite d'un item lexical. Le redoublement par écrit a des avantages pédagogiques aussi parce que la façon dont un item lexical est appris (sa *situation*) a une influence sur son rappel (Nation et Webb, 2011). Selon ces auteurs, la situation inclut la modalité (oral/écrit), le type de savoirs requis (réceptif/productif), l'endroit où le mot a été appris, et les participants qui étaient présents. Quand un mot est répété par écrit, l'apprenti-enseignant facilite le rappel futur en augmentant les modalités d'apprentissages et par conséquent les situations possibles dans lesquelles le mot pourra être rappelé ultérieurement. Enfin, la reconnaissance de mots en L2 est plus facile par l'écrit que par l'oral. Par une tâche de décision lexicale, où des sujets devaient dire si une suite de lettres ou de sons constituait un

vrai mot ou un mot inventé, Hilton (2003) a démontré que les non-natifs ont plus de facilité à reconnaître la forme écrite d'un mot que la forme orale.

Dans l'exemple suivant, Mélissa répète par clavardage la forme d'un item lexical dont l'apprenante vient de demander une traduction française.

Transcription de la parole		
1	ANA	euh parce que: je ne sais pas euh si: vous avez: (0,3) comment est-ce qu'on dit "noticed"
2	MEL	euh: remarqué
3	ANA	si vous avez: remarqué/
4	MEL	rem- (0,4) mh mh/ <((clavardage)) remarquer>
5	ANA	oui (.) remarqué/

Exemple 7 : Renforcement de la forme d'un item lexical par clavardage

L'utilisation du clavardage pour se focaliser sur la forme des items lexicaux se résume comme suit :

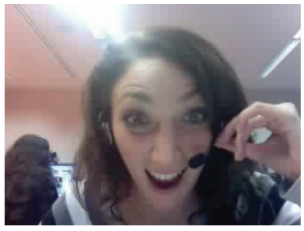
Adèle	Mélissa	Victor	Samia	Étienne	Séverine	Emilie	Total
33	22	14	15	0	8	6	98

Figure 4 : Utilisation du clavardage pour se focaliser sur la forme des items lexicaux

Le clavardage a été utilisé dans près d'un tiers des séquences d'explication lexicale. Les apprentis-enseignants semblent donc être à l'aise avec cette ressource.

6.2.2.2.2 Expressions faciales

Les expressions faciales, mouvements de tête et gestes manuels sont utilisés par les apprentis-enseignants pour se focaliser sur la forme dans une séquence sur six. Stam & McCafferty (2008 : 8) rappellent que selon McNeill (1992), « les gestes surviennent avec des éléments ayant un *dynamisme communicatif*, en d'autres mots, des informations nouvelles, focalisées ou contrastives » (notre traduction)[6]. Dans l'exemple suivant, Adèle utilise le logonyme « parler » et son visage pour rendre saillant l'item lexical :

Parole orale			Image d'Adèle	Phénomènes observés
1	ADL	et si on parlait/ euh donc on a parlé du travail/ si on parlait des CONGÉS\ (1,6)		Expression surexcitée
2	ALN	des congés/ **(0,5)**		
3	ADL	mh mh/		

Exemple 8 : Haussement de sourcils pour marquer un item lexical

[6] "Gestures occur with elements that have high *communicative dynamism*, in other words, new, focused, or contrastive information" (Stam & McCafferty, 2008: 8).

Adèle introduit le sujet de discussion en utilisant le logonyme « parler » et en levant sa voix sur la prononciation du mot « congés » (ligne 1). Tout de suite après, elle maintient un visage surexcité (yeux grands ouverts, bouche ouverte, grand sourire, sourcils levés) afin de capter l'attention des apprenantes. Cette stratégie semble avoir atteint son objectif car Alannah, une des apprenantes, demande « les congés ? » immédiatement après (ligne 2). Les items lexicaux ont été marqués par des haussements de sourcils dans 13 séquences.

Une autre expression faciale utilisée pour marquer les items lexicaux est celle de l'incertitude. C'est ce que fait Adèle dans l'exemple 14 ci-dessous quand elle donne la forme du mot « thyme ».

6.2.2.2.3 Regards

Un apprenti-enseignant peut rendre saillante la forme d'un item lexical en regardant directement vers la caméra lors de l'énonciation. C'est ce que fait Mélissa dans l'exemple 28 ci-dessous où elle donne la forme « employé ».

6.2.2.2.4 Mouvements de tête

Certains apprentis-enseignants utilisent des mouvements de tête ou des changements de postures, par exemple des rapprochements de l'écran, pour rendre saillante la forme d'un item lexical. Dans l'exemple suivant, Samia bascule sa tête de droite à gauche lors de l'énonciation de deux nouvelles formulations qu'elle apprend à Aiden, son apprenant.

Parole orale	**Image d'Emilie**	**Phénomènes observés**
1 EMI qui travaille/ (.) comme (.) serveur/ **(0,4)**		Mouvement de tête vers la gauche de l'écran
ou/ (.) en tant que		Mouvement de tête vers la droite de l'écran

Exemple 9 : Mouvements de tête pour marquer des items lexicaux

Cet exemple montre que lors des échanges par visioconférence, la tête peut devenir une ressource communicative pour marquer des items lexicaux.

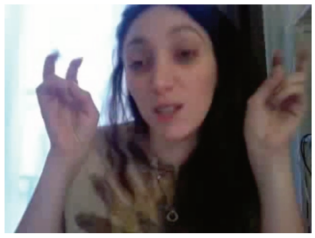
6.2.2.2.5 Les mouvements de bouche exagérés

Adèle semble être consciente de la visibilité de sa bouche, car à neuf occasions, elle rend saillante la forme d'un item lexical en faisant des mouvements de bouche grands et exagérés. C'est ce que fait Adèle dans l'exemple 22 ci-dessous quand elle corrige la prononciation. Elle use aussi de cette stratégie quand elle corrige le genre du mot « vent » en ouvrant sa bouche devant la caméra en prononçant « la ».

6.2.2.2.6 Gestes

Une façon d'introduire un autonyme est d'utiliser des guillemets. Rey-Debove (1997 : 70) estime que Frege (1893) fut le premier à utiliser des guillemets simples (et non les « guillemets ordinaires qui annoncent des paroles rapportées, des citations du langage naturel ») pour marquer le statut autonymique du langage. Au fil de l'histoire, d'autres signes ont été adoptés pour signaler le statut autonymique du langage comme des demi-crochets hauts. Parfois l'autonymie n'est pas marqué du tout. Les guillemets sont favorisés, selon l'auteure, après les verbes de signification, et l'italique est favorisé après les verbes de dénomination. Il y a tout de même trois types de marqueurs qui ne signifient pas l'autonymie : (1) « les guillemets ou l'italique de la connotation autonymique parfois appelés, en anglais, scare-quotes » (72) ; (2) les italiques de l'emphase ; (3) « l'italique de l'expression d'origine étrangère, spécialement latine, mais appartenant au lexique l1 de la langue L1 du discours » (72) par exemple *a priori*.

Dans l'exemple suivant, Adèle utilise le logonyme « on dit » et fait des gestes pour mettre l'emphase sur l'item lexical qu'elle en train d'expliquer.

Parole orale	**Image Participant**	**Phénomènes observés**
1 ADL mais en france on dit/ hum\ (0,3) pour dire hipsters/ (0,4) **on dit/ bobo**		Adèle fait des guillemets gestuels

Exemple 10 : Adèle fait des guillemets gestuels

L'utilisation de la webcam pour se focaliser sur la forme se résume comme suit :

	Adèle	Mélissa	Victor	Samia	Étienne	Séverine	Emilie	Total
Expression d'incertitude	9	2			1			12
Haussement de sourcils	12						1	13
Regard caméra		5						
Mouvements de tête	9	1	1			1	2	14
Mouvements de bouche exagérés	8							
Gestes manuelles	4	5	2		1			12
Total	41	13	3	0	2	1	3	53

Figure 5 : Utilisation de la webcam pour se focaliser sur la forme

Comme on le voit, le microphone, le clavardage et la webcam peuvent être employés afin de mettre en exergue la forme d'un item lexical. Dans la section suivante, nous allons dégager des fonctions pédagogiques de cette focalisation sur la forme lors des séquences d'explication lexicale.

6.3 Fonctions didactiques de la focalisation sur la forme

Nous avons décrit plusieurs outils à disposition des apprentis-enseignants pour attirer l'attention de l'apprenant vers la forme d'un item lexical. Dans cette section, nous allons analyser la façon dont les apprentis-enseignants les mettent en œuvre afin de faciliter l'apprentissage du vocabulaire. Dans un premier temps, nous allons analyser la focalisation sur la forme des items lexicaux lors des problèmes réels ou anticipés de compréhension. Dans un second temps, nous allons regarder la focalisation sur la forme lors des problèmes de production.

6.3.1 Problèmes de compréhension

La focalisation sur la forme des items lexicaux est nécessaire lors des séquences d'explication, surtout lors de la phase de l'ouverture où l'élément incompris est repéré, problématisé et décontextualisé. La focalisation sur la forme lors du noyau de la séquence

peut même servir à résoudre un problème d'incompréhension sans aucune explication. En aval des explications, la focalisation sur la forme aide à relier un sens à une forme afin de solidifier de nouvelles connexions dans le lexique mental des apprenants. Nous allons analyser la focalisation sur la forme à ces trois niveaux.

6.3.1.1 Vérifier la compréhension : une incompréhension anticipée

Par différents moyens, les apprentis-enseignants peuvent mettre en exergue la forme de l'item lexical dont ils veulent vérifier la compréhension. Ils le font souvent quand ils anticipent une incompréhension afin de guider l'attention de l'apprenant vers l'item lexical. Ainsi, l'apprenant peut effectuer un rappel, c'est-à-dire chercher le sens d'un item lexical dont la forme est donnée. Chaque rappel renforce la mémorisation de l'item lexical (cf. chapitre 2).

6.3.1.1.1 Vérifier la compréhension par autonymie linguistique

Une façon d'isoler et de vérifier la compréhension d'un item lexical est l'autonymie. Dans cet exemple, Séverine recourt à l'autonymie pour vérifier la compréhension du mot « systématiquement ».

Transcription de la parole	**Image de Séverine**	**Phénomènes observés**
1 SEV **et tu as (.) un employé/ (1,2) qui arrive (.) systématiquement (.) en retard\ (1,3)**		La plupart du visage de l'apprentie-enseignante est hors champ
est-ce que tu comprends/ systématiquement\ (.) naomi		Séverine s'éloigne de la webcam et lève la tête, rendant ainsi visible son visage lors de la prononciation du mot sur lequel elle se focalise

Exemple 11 : Autonymie pour mettre en exergue la forme

Dans cet exemple, le mot « systématique » obtient le statut autonymique. Quand elle dit « est-ce que tu comprends /systématique/ », /systématique/ a le statut autonymique car il signifie « mot systématique ». Ce qui est intéressant dans cet exemple est que Séverine lève sa tête et rend son visage visible lorsqu'elle prononce le mot dont elle vérifie la compréhension. Jusque-là, son visage était caché comme on le voit sur la première image.

6.3.1.1.2 Vérifier la compréhension avec l'aide du clavardage

Dans l'exemple suivant, Victor utilise le clavardage pour augmenter la saillance multimodale d'un item lexical lors d'une vérification de compréhension classique.

Transcription de la parole	
1 VIC	vous avez des pourboires/ (2,1) tu sais ce que c'est pourboire/ (1,7) <((clavardage)) pourboire>
2 CON	°pourboires:° (1,9) non

Exemple 12 : Vérification de compréhension avec renforcement par clavardage

Victor pose une question en utilisant un mot qui est potentiellement inconnu à l'apprenant, puis il vérifie la compréhension. Il met en exergue la forme orale en la répétant avec une dislocation à droite et rend saillante la forme écrite en l'écrivant dans la fenêtre de clavardage. Nous assistons ici à ce que nous appelons un *pléonasme multimodal* car les deux modes utilisées par Victor (sa voix et le clavardage) sont employés pour représenter le même item lexical qui fait l'objet d'une vérification de compréhension. Le pléonasme multimodal aide à guider l'attention de l'apprenant vers la forme de l'item lexical en question. Cette stratégie est souvent efficace car elle peut éviter qu'une explication ne soit donnée quand elle n'est pas nécessaire. Comme on voit dans l'exemple 17 ci-dessous, le redoublement de la forme par clavardage peut lui-même régler l'incompréhension.

6.3.1.1.3 Vérifier la compréhension implicitement par clavardage

Le clavardage peut renforcer la saillance de la forme d'un item lexical lors d'une vérification de compréhension, comme on l'a vu dans l'exemple précédent, mais il peut aussi être le seul indicateur qu'une vérification de compréhension est en cours. Dans l'exemple précédent, on aurait assisté à une vérification de compréhension même si le clavardage n'avait pas été utilisé. Dans l'exemple suivant, par contre, c'est le clavardage qui fait que l'on a affaire à une vérification de compréhension. Pendant que Samia explique l'activité pédagogique, elle utilise le clavardage pour vérifier de manière peu intrusive la compréhension.

Transcription de la parole	
1 SAM	et le but les buts c'est d'essayer de convaincre/ (2,1) euh (0,3) son: son client de (0,2) que: qu'il va y gagner chose\ (1,5) <((clavardage)) convaincre> (0,6) d'accord/ (1,6)
2 ANG	d'accord\

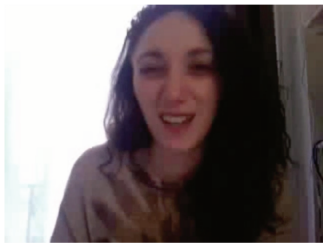
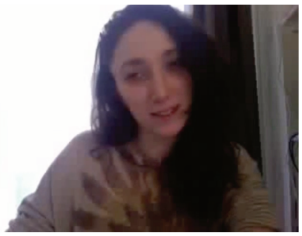
Exemple 13 : Vérification peu intrusive de la compréhension

Samia fait une longue pause de plus de deux secondes suite à la prononciation du mot « convaincre » et l'écrit dans la fenêtre de clavardage. Au lieu de vérifier explicitement la

compréhension comme l'a fait Victor dans l'exemple précédent, elle dirige implicitement l'attention des apprenants vers le mot dont elle vérifie la compréhension. En même temps, elle facilite la segmentation de la chaine parlée en isolant l'item lexical dont il est question à travers l'usage d'une autre modalité. Samia donne la possibilité aux apprenants de se manifester au cas d'une incompréhension, mais n'interrompt pas l'activité pour une vérification de compréhension explicite. Cette technique demande que les apprenants soient attentifs et se sentent à l'aise en indiquant des problèmes de compréhension. Contrairement à l'acte d'écrire au tableau dans une salle de classe, taper sur le clavier se fait de manière assez discrète. Avant l'envoi du message, le bruit du clavier et les regards vers le bas sont les seuls indices que l'apprentie-enseignante est en train d'écrire.

6.3.1.1.4 Une vérification de compréhension trimodale

Dans l'exemple suivant, Adèle utilise trois ressources multimodales à sa disposition pour problématiser un item lexical dont elle anticipe l'incompréhension. Le sujet de discussion porte sur les boissons que l'on boit quand on est malade, et en parlant du thé, Adèle explique sa méthode préférée pour créer des infusions. Au milieu d'une séquence d'explication lexicale sur le mot « infusions », elle anticipe l'incompréhension du mot « thym » et s'arrête pour se focaliser sur la forme.

Transcription de la parole		**Image d'Adèle**	**Phénomènes observés**
1 ADL	tu prends du **thym (0,5)** thym/ (0,3) euh: (1,3) thym/ (0,5) <((clavardage)) thym>		Expression d'incertitude
2 ALN 3 CTR	**(1,2)** **"thyme yeah"** (0,3) "thyme"		Tête inclinée en attendant une réponse des apprenantes

Exemple 14 : Incompréhension anticipée

Adèle focalise l'attention sur le mot « thym » en le répétant trois fois à l'oral et en l'écrivant dans la fenêtre de clavardage (ligne 1). Elle utilise une expression faciale pour montrer qu'elle anticipe la difficulté que l'item lexical peut poser aux apprenantes (ligne 1). Ces trois

modalités — la voix, le clavardage, et son visage, se complètent pour vérifier la compréhension. Elle remplit la pause de 1,2 secondes d'une nouvelle expression faciale, cette fois-ci une tête inclinée, pour montrer aux apprenantes qu'il s'agit d'une vérification de compréhension à laquelle elle souhaite une réponse. Ses apprenantes comprennent et répondent positivement en répétant le mot en anglais (lignes 2 et 3). Dans ces deux derniers exemples, on assiste à des vérifications de compréhension multimodales et efficaces où les apprenties-enseignantes font l'économie de la phrase « est-ce que vous comprenez ce que cela veut dire ? ».

6.3.1.2 Vérifier la compréhension : une incompréhension manifestée

Nous venons de voir quatre exemples de vérifications de compréhension d'items lexicaux dont l'incompréhension est anticipée par l'apprenti-enseignant. Par divers moyens, les apprentis-enseignants mettent en évidence la forme orale et/ou écrite de l'item lexical afin de faire comprendre à l'apprenant qu'il s'agit d'une vérification de compréhension. Dans cette section, nous allons analyser quelques stratégies pédagogiques par lesquelles les apprentis-enseignants se focalisent sur la forme d'un item lexical lorsque l'incompréhension est manifestée par l'apprenant au lieu d'être anticipée par l'apprenti-enseignant.

Parfois, quand l'apprenant rencontre un problème de compréhension, il ne communique pas à l'apprenti-enseignant quel élément de l'énoncé a provoqué l'incompréhension. Comme le rappelle Miecznikowski (2005), le déclencheur d'une séquence d'incompréhension peut être vague ou souligner ce qui est à réparer. Suite à un indicateur d'incompréhension, au lieu de répéter l'énoncé précédent sans modification, ce que font souvent les natifs lors de la conversation exolingue (Smith, 2003), une stratégie pédagogique utile consiste à s'efforcer d'identifier la source du problème (Schegloff, Jefferson, & Sacks, 1977) par une vérification de compréhension (Varonis & Gass, 1985). Nous allons analyser deux exemples pour mettre en contraste les effets que peuvent avoir d'un côté la focalisation sur la forme et de l'autre côté l'absence de focalisation sur la forme lorsque qu'un problème de compréhension est manifesté par l'apprenant.

6.3.1.2.1 La source d'incompréhension n'est pas identifiée

Dans l'exemple suivant, l'attention à la compréhensibilité du langage diminue lors de la clôture de la séance quand Samia demande à son apprenante ce qu'elle a de prévu pour le week-end. En sortant de son rôle d'enseignante, Samia augmente fortement le débit de sa parole, rendant incompréhensible sa question.

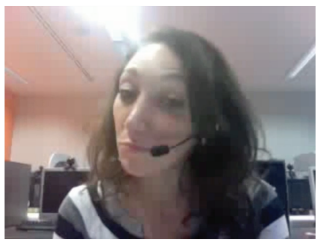
Transcription de la parole	
1 SAM	donc ça fait six minutes/ (0,4) et puis euh: (1,2) <((débit très élevé)) et puis d'accord donc t'as prévu quoi ce week-end alors> (1,4)
2 ANG	[euh: comment/]
3 SAM	<((débit élevé))[comme on a fini l]a séance (0,4) on a fini la séance/ donc on va:\ (0,7) on va: on va: parler d'autres choses/ ((rire)) t'as prévu [quoi ce week]end ((rires))>
4 ANG	[d'accord] (1,8) hum: (1,2) ce que j'ai: cuisiné/ ou euh:\ (0,5)
5 SAM	<((débit lent)) t'as PRÉVU (.) est-c- (.) qu'est-ce que tu veux faire ce week-end> (1,5)
6 ANG	oh (.) euh ce weekend/ (0,7) euh: j'irai à: glendalough/

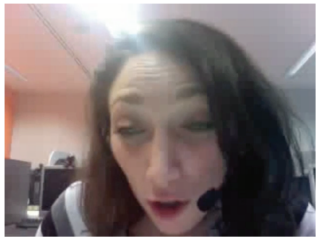
Exemple 15 : Changement de rôle d'amie à professeure

Au lieu de chercher à cibler l'élément non compris, Samia répète toute sa question. Ce n'est qu'après deux indicateurs d'incompréhension (lignes 2 et 4) que Samia se remet en position d'enseignante, réduit le débit de sa parole et rend saillant le mot « prévu » qui est probablement la source du trouble (ligne 5) (Schegloff et al., 1977). Elle décide de reformuler sa question au lieu de s'attarder sur un mot potentiellement inconnu. Par contraste, dans l'exemple suivant, Adèle, qui est dans son rôle de professeur, a le reflexe de chercher tout de suite à identifier la source précise de l'incompréhension suite à un marqueur d'incompréhension.

6.3.1.2.2 La source d'incompréhension est identifiée

Lors d'une discussion sur la pause café dans les entreprises en France, Adèle s'intéresse aux habitudes des Irlandais et utilise le mot « équivalent » qui n'est pas reconnu par l'une de ses apprenantes.

Parole orale		Image d'Adèle	Phénomènes observés
1 ADL	est-ce qu'il y a un équivalent/ (.) euh (.) en irlande/ (0,2)		Adèle tourne la tête après avoir posé une question
2 ALN	**euh est-ce que tu** répéter s'il te plaît/ (0,3)		
3 ADL	oui\ (0,3)		

	est-ce que vous connaissez le mot équivalent/ **(0,7)**		Rapprochement de l'écran, visage d'incertitude et secouement de tête lors de la prononciation de « non » (ligne 5)
5 CTR	**°non°**		
5 ADL	**[non/]**		
4 ALN	**[hum] (0,5) équi- OUI** **(0,7)**		
5 ADL	**oui/ (0,5)** **donc est-[ce qu'il (y)]&**		
6 ALN	**["yeah"]**		
7 ADL	**&toi tu ne comprends pas catriona/** **(0,2)**		
8 CTR	**non** **(0,6)**		
9 ADL	non\		

Exemple 16 : Identifier la source de l'incompréhension

Au lieu de répéter sa question, Adèle cherche à identifier la source de l'incompréhension. Elle suppose que c'est l'item lexical « équivalent » et vérifie son hypothèse en mettant en exergue la forme orale de celui-ci en utilisant le logonyme « mot » (ligne 3). Grâce à cette focalisation sur la forme, elle peut désormais consacrer son énergie à la la résolution du problème, rendant ainsi son enseignement plus efficace et économe.

6.3.1.3 Résoudre rapidement l'incompréhension d'un mot transparent

Parfois, une explication par définition, traduction ou images n'est pas nécessaire. Si un mot est transparent entre le français et la langue maternelle (ou une autre langue) des apprenants, la focalisation sur la forme peut suffire pour résoudre l'incompréhension et relier une forme à un sens. Observons la suite de la séquence d'explication lexicale dont l'ouverture est représentée dans l'exemple précédent.

Transcription de la parole	
10 ADL	alors un équivalent/ je te l'écris\ (1,1) <((clavardage)) un équivalent> un équivalent/ (0,8) ça veut dire/ (.)[euh:\]
11 CTR	[OH\]
12 ADL	[tu connais/]
13 CTR	[ah oui\] (0,4) oui m- (.) oui\ (0,7)
14 ADL	voilà\ alors est-ce que en irlande/ (0,4) il y a/ un équivalent/ (.) de la pause (0,3) café/ (0,5) (que) vous faites aussi dans les entreprises/

Exemple 17 : Résoudre l'incompréhension en mettant en exergue la forme

La focalisation sur la forme écrite du mot transparent, accompagnée d'un commentaire méta-sémiotique (« je te l'écris », ligne 10) permet à Catriona de comprendre tout de suite le sens du mot qui a une forme quasiment identique en anglais (*equivalent*). Dans cette séquence d'explication, la focalisation sur le sens a une double fonction d'identifier la source

d'incompréhension dans un premier temps est de la résoudre rapidement dans un second temps.

Les exemples précédents montrent l'utilité pédagogique de mettre en exergue la forme d'un item lexical lors des ouvertures des séquences d'explication lexicale. Si l'incompréhension est anticipée par l'apprenti-enseignant, la focalisation sur la forme aide l'apprenti-enseignant à vérifier la compréhension afin de juger la nécessité d'une explication. Si l'apprenant manifeste son incompréhension et ne donne pas un indicateur qui identifie l'élément langagier non compris, la focalisation sur la forme d'un item lexical est d'une grande utilité à l'apprenti-enseignant car elle lui permet de cibler l'élément non compris, qui est potentiellement un item lexical, et de mieux rentabiliser les 40 minutes de la séance. Si, à cause d'un problème de décodage phonologique, l'apprenant croit par erreur ne pas connaître un item lexical, la focalisation sur la forme (surtout écrite) de celui-ci par l'apprenti-enseignant peut résoudre rapidement l'incompréhension et faire l'économie d'une explication. La focalisation sur la forme peut être utile aussi après une explication afin de solidifier le lien entre le sens et sa forme. Les exemples qui suivent vont explorer ce phénomène.

6.3.1.4 Relier la forme au sens

Dans les exemples suivants, les apprentis-enseignants se focalisent sur la forme d'un item lexical afin de la relier à une définition qui précède, que la définition vienne de l'apprenti-enseignant ou de l'apprenant. L'exemple suivant est issu d'une mise en situation dans laquelle les apprenants, jouant le rôle d'un manager dans une entreprise, doivent réagir à plusieurs types d'employés problématiques. L'incompréhension du mot « paresseux » est anticipée par Étienne qui le définit.

Transcription de la parole	**Image d'Étienne**	**Phénomènes observés**
1 ETN maint`nant/ si: il y a un employé qui: ne travaille pas assez il est (.) on dit qu'il est **paresseux**\ (0,7) qu'est-ce que vous lui dites (1,2) 2 SOH °euh° (0,5) euh: (0,6) pouvez: répéter s'il vous plait/ (0,8)		Battement vers le bas de l'écran

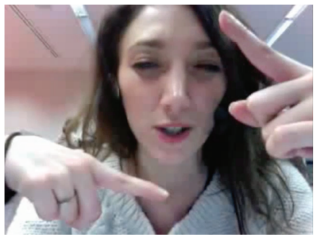
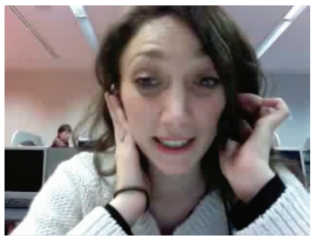
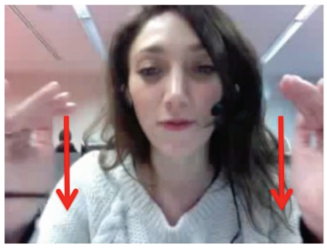
3 ETN 4 SOH 5 ETN 6 SOH 7 ETN	un employé qui est paresseux/ **(0,4)** paress- ça veut dire qu'il ne travaille pas assez/ ah\ (0,6) euh: (0,4) qu'est-ce que vous lui dites\		Battement hors champ

Exemple 18 : Introduire un mot qui vient d'être défini

Après avoir défini le mot « paresseux », il souligne sa forme afin de la relier au sens qu'il vient de donner (ligne 1). Pour ce faire, il emploie le logonyme « on dit » et fait un battement avec sa main droite lors de la première prononciation du mot. Ce qui est intéressant ici est le fait qu'il refait le même battement deux fois, chaque fois synchronisé avec le mot « paresseux ». Il semble ne pas être conscient du champ de sa caméra, car son deuxième geste est à peine visible.

Cette explication est l'une parmi plusieurs où la définition est donnée avant la forme orale ou écrite. Ceci augmente la charge mémorielle imposée à l'apprenante car elle doit se souvenir de la définition précédente et la relier à la forme qui n'est donnée qu'après. Dans cet exemple, la première définition semble ne pas avoir été efficace, car Sophie demande une répétition à la ligne 2. Lors de sa deuxième explication (ligne 3), Étienne procède à l'ordre inverse en mettant en exergue la forme du mot par (par un battement et une pause) avant de donner une définition. Sophie est ainsi prête à écouter la définition (ligne 5) et semble avoir compris (ligne 6).

Souvent, les apprentis-enseignants saisissent des moments où les apprenants parlent de leur vie personnelle, scolaire ou professionnelle afin de privilégier l'apprentissage incident du lexique. Lors de ces séquences non-planifiées, les apprentis-enseignants donnent des mots ou expressions aux apprenants qui résument de manière succincte ce que les apprenants viennent de dire en plusieurs phrases, ou donnent des items lexicaux qui sont liés au sujet. Dans l'exemple suivant, Adèle donne l'expression « faire le tri » afin d'enrichir à la fois la discussion sur le CV et les connaissances lexicales de l'apprenante.

Parole orale		Image d'Adèle	Phénomènes observés
1 ADL	et hum: est-ce que ça vous: a semblé: hu:m (0,6) plus d- (.) plus difficile/ ou pas de faire (.) **une seule page (0,4)** c'est MIEUX/ ou c'est moins\ (0,9)		Geste iconique représentant « une seule page »
2 ALN	c'est difficile pour nous par- (.) euh: (0,4) car euh: (1,4) euh: tout euh mes c v euh: mon c v est déjà (0,6) deux pages/ (0,9) euh en anglais/ "so" (0,5) euh: (0,3) il a fallu qu- euh qu'on qu'on fasse euh: un c v euh plus (.) plus court/ (0,6) court/ (0,9) euh et plus con-		
3 ADL	faire le TRI/		
4 ALN	euh: pardon/ (1,4)		
5 ADL	l'expression (.) **FAIRE LE TRI/** (1,1)		Rapprochement de l'écran et haussement de sourcils
6 ALN	euh x		
7 ADL	ça veut [dire]		
8 ALN	[xx/] (0,9)		
9 ADL	ça veut dire/ (.) euh: <((clavardage)) faire le tri> SÉLECTIONNER/ (.) **les informations [importantes/]**		Paumes orientées vers le bas, mouvement descendant des deux mains ensemble
10 ALN	[ah\ (.) oui] (0,5) euh: (1,2) x (1,6) [mh:]		
11 ADL	[donc vous avez] fait le tri/ (0,9)		
12 ALN	et euh: c'était euh plus concis		

Exemple 19 : Relier la forme de l'item lexical à son sens

Dans le cadre de leur module d'apprentissage du français, les apprenants doivent élaborer un CV en français d'une seule page. Pour leur en faire parler, Adèle leur demande s'il est difficile de ne faire qu'une page. Alannah à la ligne 2 explique la difficulté qu'elle a rencontrée en devant raccourcir son CV de deux pages qu'elle avait déjà crée. Adèle profite de cette opportunité pour donner à la ligne 3 l'expression « faire le tri », qui est à la fois un régulateur d'écoute et un item lexical inconnu à l'apprenant. Il devient l'objet d'une séquence d'explication et Adèle le nomme ainsi à la ligne 5, quand elle utilise le logonyme « expression », hausse les sourcils, introduit une définition en disant « ça veut dire » (ligne 7)

et le redouble à l'écrit (ligne 9). Sa phrase à la ligne 11 est à la fois une façon de relier la forme à la définition suite à celle-ci, et de clôturer la séance de travail sur la langue afin de retourner vers la conversation interrompue.

Dans l'exemple précédent, la forme d'un item lexical a été reliée à une définition qui a été élaborée à la fois par l'apprenante et par l'apprentie-enseignante. Dans l'exemple qui suit, l'apprentie-enseignante donne la forme d'un item lexical dont le sens a été élaboré uniquement par l'apprenante. Alannah vient de parler d'un permis de conduire qui se fait par ordinateur et qui atteste des compétences acquises. Adèle donne le nom en français en utilisant le clavardage.

Transcription de la parole	
1 ADL	ah: (1,1) en france on appelle ça comme ça\
	<((clavardage)) tice>
	(0,8)
2 ALN	"a tice well" c'est l- [c'est] le [même]
3 ADL	[(°si°)] (0,4) [c'est ça/]
	(0,5)
4 ALN	j[e crois que]
5 ADL	[ah ben c'est ça\ o]k

Exemple 20 : Donner la forme d'un item lexical dont le sens a été expliqué par l'apprenant

Les apprenants adultes ont déjà beaucoup de connaissances sur le monde, et l'apprentissage d'une langue étrangère consiste à relier de nouvelles formes à des unités de sens déjà développées en langue maternelle. Dans l'exemple précédent, Alannah n'a pas à apprendre le concept de TICE car elle sait déjà ce que c'est. Elle n'a qu'à apprendre la forme en français, qui est en l'occurrence pareille que la forme en anglais, or alors elle pense que c'est pareil. La focalisation sur la forme aide à établir ces nouvelles connexions dans le lexique mental (Bogaards, 1994).

6.3.2 Synthèse : problèmes de compréhension

Dans cette section, nous avons analysé différents moyens par lesquels les apprentis-enseignants mettent en exergue la forme orale des items lexicaux qui font l'objet de séquences d'explication lexicale, ainsi que les fonctions pédagogiques de cette focalisation sur la forme. Nous avons repéré quatre fonctions pédagogiques de la focalisation sur la forme lors des problèmes de compréhension. La focalisation sur la forme :

1. Assure qu'il y ait orientation mutuelle vers un item lexical qui va faire par la suite l'objet d'une explication. Plusieurs outils multimodaux sont à disposition des apprentis-enseignants pour rendre saillante la forme d'un item lexical.

2. Permet aux apprentis-enseignants de vérifier de manière explicite ou implicite la compréhension d'items lexicaux qu'ils jugent utiles ou potentiellement problématique afin de juger la nécessité d'une explication. C'est particulièrement utile dans une interaction qui ne dure que 40 minutes une fois par semaine. Lorsqu'un apprenant rencontre un problème de compréhension, il y a de fortes chances que l'item inconnu soit de type lexical (Nicolaev, 2010; Smith, 2003). Puisque les apprenants n'ont pas toujours le reflexe de souligner eux-mêmes le déclencheur (l'élément non compris), la focalisation sur la forme permet à l'apprenti-enseignant d'identifier l'élément qui a besoin d'être expliqué par la suite.
3. Peut résoudre rapidement l'incompréhension d'un item lexical partiellement connu, surtout pour les mots transparents, si la source de l'incompréhension est le décodage phonologique qui est souvent problématique en langue étrangère.
4. Sert à relier la forme à une définition qui précède.

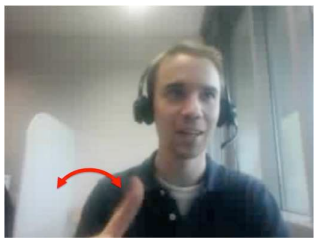
6.3.3 Problèmes de production

Dans la section précédente, nous avons investigué la focalisation sur la forme et ses fonctions pédagogiques lors des problèmes de compréhension. Ici, nous allons analyser la focalisation sur la forme lors des problèmes de production. Pour rappel, un problème de production est un moment où un apprenant subit une panne lexicale (Merlino & Traverso, 2009) lors d'une tâche productive. Pour résoudre la panne, l'apprenti-enseignant doit donner à l'apprenant le *mot juste* (Grossmann, 2011).

6.3.3.1 Travailler la prononciation

Quand un apprenant rencontre un problème de prononciation, le sens du mot peut être connu mais la forme n'est que partiellement connue ou complètement inconnue. La forme orale est une des facettes de la connaissance d'un mot qu'il faut maîtriser (Nation, 2013), ce qui fait que le travail sur la prononciation fait partie du travail sur le lexique. Nous allons regarder deux exemples de travail sur la prononciation, l'un initié par l'enseignant lors d'une tâche communicative et l'autre initié par l'apprenante lors d'une séquence de travail sur la langue.

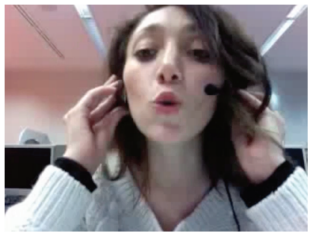
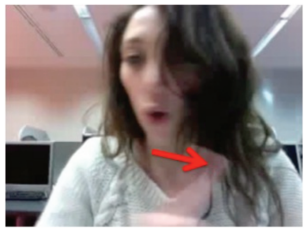
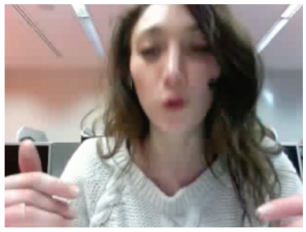
Dans ce premier exemple, Victor corrige la prononciation du mot « gouvernement », qui est un mot transparent pour les anglophones, après que son apprenant Liam le prononce [guvɛrmã].

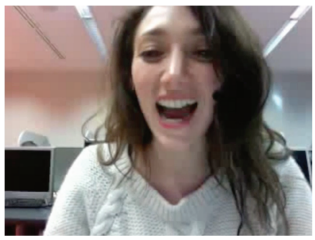
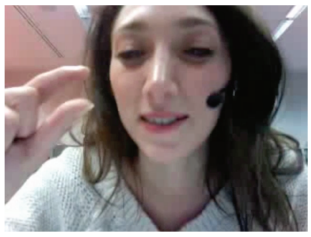
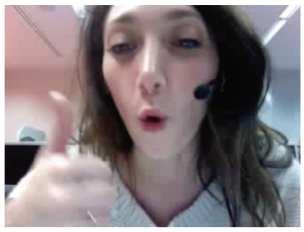
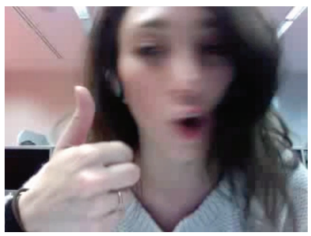
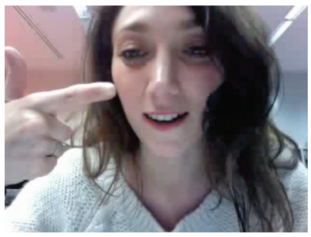
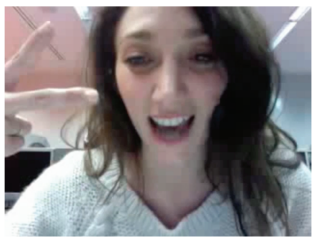
Parole orale	Image de Victor	Phénomènes observés
1 LIM comme euh le {guvɛrmã} (1,0) 2 VIC c'est vrai (0,5) 3 LIM oui (0,8) 4 VIC alors (.) en français/ (0,9) on di- on prononce/ **hum:** **(0,5) la deuxième syllabe** **c'est gouverNEment** (0,9) 5 LIM gouvernement\ (0,3) 6 VIC oui (.) très bien\ (0,4) e[xce]llent 7 LIM [°oui°]		Basculement de l'index de gauche à droite trois fois

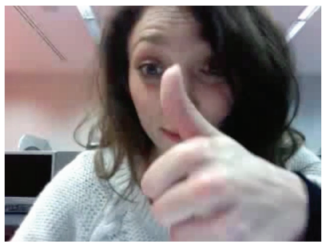
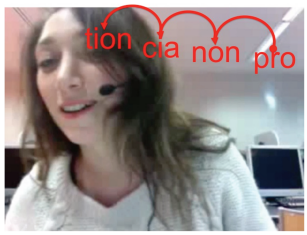
Exemple 21 : Corriger la prononciation du mot « gouvernement »

Avant de corriger la prononciation, Victor réagit sur le contenu du message de Liam, ce qui témoigne du caractère résolument communicatif de l'échange. Pour marquer le passage à une séquence de travail explicite sur la langue, Victor emploi le mot « alors » ainsi que les logonymes « en français » et « on dit » (ligne 4). Victor accompagne sa description linguistique d'un geste qui guide l'attention de l'apprenant vers la prononciation de cette syllabe qui est inexistante dans la version anglaise du mot. Cette explication s'avère efficace car Liam prononce le mot correctement à la ligne 5 et se fait évaluer positivement par l'apprenti-enseignant à la ligne 6.

Dans ce second exemple, Alannah est en train d'expliquer que le bilan créé par Adèle lui est utile car il lui permet de se réécouter et de se rendre compte de ses erreurs de prononciation, ce qui déclenche une séquence d'explication sur la forme orale du mot « prononciation ».

Parole orale	Image d'Adèle	Phénomènes observés
1 ALN euh par exemp`e/ la pronon-/ (0,4) pronona- (0,2) (pon-) (0,5) pronona- (1,5) 2 ADL **PRONONCIATION** 3 ALN "yeah" (.) le prononnonci- (0,5) 4 ADL (cia) (0,3) ((rire)) 5 ALN ((rire)) 6 ADL c'est dur à dire\ (0,2) [PR]O (.) [NON] (.) CIA (.)& 7 ALN [le] (0,4) [le] 8 ADL &TION\ (0,8) 9 ALN le pronona- (0,4) 10 ADL **PRONONciation** (0,7) [((rire))] 11 ALN [prononna: (0,6) ((phrase incompréhensible en anglais)) (0,3) pro:non: (0,8) na:tion] (0,3) 12 ADL c'est [MIGNON même comme ç]a\ 13 ALN [pro (.) pronona-] (0,5) prononation (0,4) euh [((rire))] 14 ADL **pron-** (0,1) [non\ (.) non\ (.) attends-tends\ (0,3) regarde reg- ((rire))] 15 ALN oh: (.) merde\ (0,6) ((rire)) (0,2)		Tête levée lors de la prononciation
16 ADL PRO/ <((clavardage)) pro> (1,4) 17 ALN pro/ (1,6) 18 ADL NON/ <((clavardage)) non> (0,4) 19 ALN non/ (0,5) nation\ (0,3) 20 ADL comme **non/** (.) pro (.) **non/** (0,4) 21 ALN [pro]nona (.) [si]- (.) pro & 22 ADL [nan] <((clavardage)) cia>	 	Index tendu lors de la prononciation de « non » dans « comme **non/** » Deux mains levées lors de la prononciation de « non » dans « pro (.) **non/** »

(0,5) [**CIA**] (0,5)& 23 ALN & (.) n[on] 24 ADL & [**CIA**] (0,6) 25 ALN cia (0,7) 27 ADL [**CIA**/] 28 ALN [pronon]ation 29 ADL CIA\ (1,2) 30 ALN cia\ (1,0)		Grand mouvement de bouche lors des trois premières prononciations de « **CIA** »
31 ADL **dis (.) dis une par une** (0,4)		Geste métaphorique qu'elle dirige vers le bas en disant « une par une »
pro/ **(0,5)** 32 ALN pro/ **(1,5)**		Pouce levé et maintenu
33 ADL **non/**		Secouement de tête
(0,5) **34 ALN non/** **(0,2)**		Index tendu et maintenu
35 ADL **cia/** **(0,5)** 36 ALN **cia/ (.) non** (1,5) [tion\] 37 ADL <((clavardage)) tion> [MH\] (0,2)		Hochement de tête puis doigts tendus et maintenus lors de la prononciation de la troisième syllabe

MH\ **(0,3)** [TION\] 38 ALN [pronon] (0,3) prononation\ (1,3) prononc- [((rire))] "oh: g[o:d"] 39 ADL [bien/] (1,2) [essaye de le dire] doucement 41 ALN ((rire))		Haussement de sourcils et pouce levé
42 ADL et dis-le tout douc`ment\ (0,3) **PRO/ (.) NON/ (0,2) CIA/ (0,2) TION** (0,2) 43 ALN pronona:tion\ (1,4) ah 44 ADL oui/ (0,5) c'est MIEUX qu'au début déjà c'est bien 45 ALN [j'espère (.) ((rire))] 46 ADL [((rire)) il est dur à] dire hein/ (1,2) d[onc ça ça t'aide/ ça vous aide/] 47 ALN [ah je suis f-/ ah je suis] fatiguée (c'est) (1,0) 48 ADL oui: je comprends\ vous travaillez beaucoup\ (0,4) [((rire))] 49 ALN ((rire)) (0,5) ("no")		Changement de position de la tête par syllabe, de droite à gauche de l'écran

Exemple 22 : Explication multimodale de la prononciation du mot « prononciation »

Dans cette séquence, Adèle utilise une variété de ressources multimodales à sa position pour aider l'apprenante à se former une représentation phonologique de l'item lexical « prononciation ». Elle commence par répéter la forme correcte lentement avec sa bouche levée et centrée devant la webcam (ligne 2). Après un mot d'empathie, elle répète le mot syllabe par syllabe (lignes 6-8). Quand l'apprenante est toujours incapable de le prononcer correctement, Adèle répète sa stratégie de diviser le mot en syllabes mais augmente la saillance multimodale en redoublant par écrit chacune des syllabes. En plus de ce pléonasme multimodal oral et écrit, Adèle crée un lien mental entre « non » (deuxième syllabe du mot) et « non comme non » (ligne 20) accompagné d'un geste emblématique voulant dire « non ». Elle maintient sa bouche ouverte lors de la prononciation de la troisième syllabe « cia » pour la rendre encore plus saillante. Ensuite, elle verbalise sa stratégie de séparation syllabique en disant « dis une par une » en faisant un geste métaphorique, ses deux doigts tenant des syllabes imaginaires. Elle répète encore une fois le mot « prononciation » par syllabes, en maintenant devant la webcam les gestes indiquant le numéro de syllabe (lignes 32-37). Elle félicite l'effort exercé par l'apprenante et répète une dernière fois le mot, cette fois-ci en changeant la position de sa tête pour marquer les quatre syllabes (ligne 42).

Avec patience et savoir-faire multimodal, Adèle utilise sa voix, le clavardage, son visage, ses mains, et la position de sa tête pour illustrer la forme orale du mot. Qu'elle le

sache ou pas, elle emploie aussi une partie de la technique du mot-clé (cf. chapitre 2, section 2.4.6 en page 70). Dans cette séquence, Adèle adapte cette technique pour relier la deuxième syllabe de « prononciation » au mot « non », et elle anime ce lien gestuellement par un secouement de tête et un geste emblématique (ligne 20). On peut se demander en outre si le nombre élevé de tours de parole dédiés à cette séquence, ainsi que l'absence de marqueur de passage d'une séquence communicative à une séquence de travail explicite sur la langue, sont dus au fait que les participantes étaient déjà en train de parler de la langue en parlant du bilan.

Dans les deux exemples qui précèdent, nous avons vu une explication de la bonne prononciation d'un mot transparent (gouvernement) et une animation multimodale par syllabes de la prononciation d'un mot (prononciation), adoptant certaines parties de la technique du mot-clé. Dans ces deux exemples, on observe l'usage de gestes pédagogiques (Tellier, 2008a) servant à véhiculer de l'information phonologique. Ce travail sur la prononciation augmente les compétences communicatives des apprenants. Le cas de « gouvernement » illustre le fait que l'existence d'un mot transparent ne garantit pas automatiquement qu'il soit exploitable correctement dans une situation de communication orale. Afin de « connaître » un mot, il faut connaître toutes ses facettes, y compris sa prononciation. Le travail sur « prononciation » dans le second exemple est particulièrement pertinent car c'est l'apprenante qui soulève la difficulté. Ces interactions en visioconférence sont des moments privilégiés pour une prise en compte des besoins individuels des apprenants.

6.3.3.2 Travailler l'orthographe

Bien que l'objectif de ce cours en ligne soit l'entraînement à l'oral, un travail explicite sur l'orthographe s'observe à certains moments pour diverses raisons. Dans l'exemple suivant, la forme écrite d'un item lexical est mise au premier plan par Mélissa suite à une faute d'orthographe commise par son apprenante Ana. Quand le canal audio d'Ana cesse de fonctionner, elle décrit par écrit son métier idéal dans une entreprise de traduction :

Transcription de la parole	
1 ANA	<((clavardage)) Je voudrais un entretien ou je peux utiliser mes trois langues. J'aimeraisfaire mon stage dans un entreprise de traducion> °traduction/° (0,8)
2 MEL	.tsk alors ((raclement de gorge)) (1,6) oui:/ (0,4) donc c'est quoi alors français: espagnol et euh: et anglais\ c'est ça/
3 ANA	ouais
4 MEL	oui
5 ANA	ouais

6 MEL	mh mh/ (1,0) euh: dans une entreprise/ de traduction\ (0,2) d'accord\ (0,4) traduction/ c'est avec un d\ hein/ (.) tra- (0,4) DUC/ (0,2) tion\ <((clavardage)) traduction> (0,9) hu[m:]
7 ANA	[ah okay] (0,5)
8 MEL	okay du coup: euh ouais/ (0,2) et tu voudrais faire ça plutôt: euh: (.) en france: en irlande/ euh: en espagne/

Exemple 23 : Corriger l'orthographe

Mélissa, comme Victor dans l'exemple 20 ci-dessus répond d'abord au contenu du message avant de s'attarder sur la forme. Quand Mélissa corrige l'orthographe, elle isole la syllabe problématique avec saillance vocalique et la commente avant de donner la forme écrite correcte et continuer l'activité communicative (ligne 6). L'objectif de la tâche étant communicatif, le clavardage sert ici d'outil de correction peu intrusif. Cet exemple nous rappelle qu'en visioconférence, les pannes techniques peuvent arriver à tout moment. Travail sur la langue devient alors travail sur la langue écrite.

6.3.3.3 Augmenter les compétences communicationnelles des apprenants

Parfois un apprenant a besoin de connaître la forme écrite d'un item lexical afin de communiquer plus efficacement en milieu professionnel. L'objectif de ce cours en ligne est d'améliorer les compétences orales, mais aussi de préparer les apprenants à faire un stage en France. Une des compétences requises pour cela est l'élaboration d'un CV qui est un objet écrit. Cet exemple est issu d'une conversation entre Samia et son apprenante Angela à propos des compétences que cette dernière pourrait ajouter à son CV. Samia emploie le mot « gérer » et met sa forme orale tout de suite en exergue, soit parce qu'elle anticipe un éventuel problème de compréhension, soit parce qu'elle souhaite qu'Angela le note. Suite à une explication de son utilité en tant que mot-clé sur un CV, Angela en demande la forme écrite, ce qui veut dire que la focalisation sur la forme semble avoir été une stratégie efficace pour piquer l'attention de l'apprenante.

Transcription de la parole	
1 SAM	tu as appris à: est-ce que tu as appris à: par exemple à GÉRER/ (1,5) un groupe/ (0,4) tout simplement/ (1,9)
2 ANG	oui j'espère que j'ai: (0,2) ((rire))
3 SAM	donc ça c'est important tu vois ça c'est une compétence (0,5) apprendre à euh être capable de (0,3) gérer un groupe/ c'est une compétence que tu peux mettre sur un c v\ (4,1) tu as compris/ (1,1)
4 ANG	oui (0,9)
5 SAM	par exemple pour ton c v euh que tu vas déposer sur le site euh quand t- euh (0,3) quand tu vas chercher ton stage à reims (0,9)

```
          si il faut qu`tu mettes des com- (.) il va falloir que tu mettes
          des compétences/ (1,0) tu pourras par exemple mettre que tu sais
          GÉRER/ (0,8) un groupe\ (0,3) parce que tu as déjà eu à gérer un:
          groupe d'enfants\
          (4,0)
6 ANG     euh: comment est-ce que ça s'écrit\
          (1,2)
7 SAM                       [GÉRER un groupe <((en riant)) bah>]
8 ANG                       [c'est euh:]
          (1,7)
9 SAM     et tu pourras m'envoyer si tu veux ton c v pour que j`te corrige
          euh:
          <((clavardage)) gérer un groupe>
          ou si tu as b- qu- quelle que soit ta question euh si tu en as
          une/ (0,3) tu pourras m`la poser/ (0,5) d'accord/ (0,2)
          j'[avais avais donné mon fa]cebook la la la semaine dernière/
10 ANG      [merci]
```

Exemple 24 : Donner la forme écrite d'un mot-clé

L'intérêt didactique de cette séquence réside dans le fait que l'apprenante sera, grâce à cette interaction, capable de réutiliser ce mot nouveau sur un CV français. Quand on augmente le nombre de facettes connues d'un item lexical, on augmente la capacité communicative. L'habilitation à faire un stage en France étant un des objectifs principaux de ce cours, l'attention que porte l'enseignante sur cet item lexical, en particulier sur ses usages et sur sa forme écrite, témoigne du côté pratique de ce module de formation.

6.3.3.4 La focalisation sur la forme pour distinguer des homophones

Quand un apprenant confond deux homophones, la focalisation sur la forme écrite s'impose et le clavardage devient un outil efficace. A certains moments pendant la cinquième séance, Ana rencontre des problèmes techniques et doit s'exprimer par clavardage. L'extrait suivant est issu d'une séquence de négociation pendant laquelle les participantes doivent élaborer un menu de camion-restaurant pour une clientèle dite « bobo » dans un quartier branché de Lyon.

Transcription de la parole	
1 ANA	<((clavardage)) je pense que il faut que vous fasse attencion et vous doit vous installerpres du lieu de travaille des bobos> (4,1)
2 MEL	je pense (0,4) qu'il faut: que vous <((en riant)) fassiez attention/> (1,0) vous devez (.) vous installer près du lieu où tr- ah d'accord\ (0,8) près du lieu de travail/ (1,0) alors travail c'est comme ça hein/ (0,8) <((clavardage)) travail> (0,7) sinon c'est le verbe\ (0,5) avec deux l e c'est le verbe/ (0,7) et avec (0,8) un seul l/ (0,2) c'est le nom\ (0,8) près du lieu de travail\ oui c'est une bonne idée/ (0,2) c'est vrai j`vais retenir l'idée/

Exemple 25 : Focalisation sur la forme pour aborder un point grammatical

Ana confond « travail » et « travaille », qui est une erreur qui n'aurait pas été rendue visible si elle s'était exprimée à l'oral. Comme dans l'exemple 22 ci-dessus, Mélissa saisit cette opportunité pour donner la forme correcte par clavardage. Puisque l'apprenante confond deux mots, Mélissa profite de cette opportunité pour aller plus loin et expliquer la différence entre ces deux formes qui appartiennent à la même famille de mots.

6.3.3.5 Distinguer des items lexicaux phonétiquement proches

Ce ne sont pas uniquement des homophones qui peuvent être confondus par les apprenants. Dans les deux exemples suivants, une focalisation sur la forme est utilisée pour désambiguïser des mots qui sont phonétiquement proches.

Ce premier exemple est issu d'une activité dont l'objectif est de réfléchir à la meilleure façon de gérer un employé difficile ou paresseux. Une séquence sur la prononciation est initiée par l'apprentie-enseignante quand l'apprenante, qui décrit une employée paresseuse, confond « nettoyant » et « nettoyage ».

Parole orale	**Image de Séverine**	**Phénomènes observés**
1 NAM quand: (.) nous: devons: (0,7) faire: (.) le: (.) nettoyant/ (1,1) 2 SEV **le [nettoyage]/** (0,5) [ouais] 3 NAM [e- euh] [nettoyage/] 4 SEV nettoyage/ (0,5) [ouais] 5 NAM h[u]m: 6 SEV <((clavardage)) nettoyage> 7 NAM e- elle va\ (0,5) aux: (.) toilettes (0,3) encore\ (0,4) 8 SEV d'accord\		Séverine lève la tête et se rend visible lors de la prononciation du mot juste

Exemple 26 : Mise en exergue la forme du mot « nettoyage »

Séverine répète la bonne prononciation plusieurs fois et écrit le mot juste dans la fenêtre de clavardage pour que l'apprenante en ait une représentation à la fois phonique et graphique. Ce faisant, elle lève la tête et se rend visible lors de la prononciation du mot « nettoyage », ce qui marque un changement de posture par rapport à sa position de repos habituelle où son visage reste invisible hors champ. Cette intervention semble efficace car Naomi répète la bonne prononciation à la ligne 3.

Le clavardage et la répétition à l'oral ne sont pas les seuls outils pour distinguer deux mots qui sont phonétiquement proches. Dans l'exemple suivant, Mélissa utilise la saillance vocalique pour aider l'apprenant à percevoir la différence entre « magasin » et « magazine ».

Transcription de la parole	
1 MEL	en espagne/ ils travaillent vous travaillez plutôt où/ alors/
	(0,4) °est-ce° que vous [travaillez dans des] bars/ dans des
2 ALJ	[et]
	normalement en les bars/ (.) et hum magazines/
3 MEL	ouais/ (.) [dans des m]aga[S]INS/
4 ALJ	[oh:\] [mh] (0,6) magasins/ (0,5)
	[°mh°]
5 MEL	[°puisqu]e° magazine/ tu sais c'est les catalogues\
	(0,5)
6 ALJ	AH:\ euh (0,5) [d'accord\]
7 MEL	[magazi]ne c'est les c'est les petits catalogues/
	comme les journaux/ (.) et magaSIN c'est [euh:] des boutiques
8 ALJ	[°mh°] (1,4) magasins
	(2,9)
10 MEL	et euh ouais/ (0,2)

Exemple 27 : Désambiguïsation de *magasin* et *magazine* par saillance vocalique

Mélissa commence (ligne 3) par accuser la réception de la réponse d'Alejandra, et donne le mot juste en levant sa voix lors de la prononciation de la dernière syllabe du mot « magasin » afin de le distinguer du mot « magazine ». Afin de renforcer la distinction, elle ajoute une définition à la ligne 5 afin que l'apprenante comprenne le sens du mot qu'elle a utilisé par erreur (magazine). A la ligne 7, elle répète sa définition et ajoute une définition du mot juste. Nous assistons à la désambiguïsation de deux mots phonétiquement proches grâce à l'aide de saillance vocalique sur les syllabes différentes et de courtes définitions.

Dans les exemples précédents, nous avons assisté à des séquences d'explication lexicale dans lesquelles les participants se sont focalisés sur des items lexicaux dont la forme était partiellement connue. L'objectif était de donner le bon suffixe d'un mot (Séverine, *nettoyage*), de rendre saillant les différences entre les formes anglaise et française d'un mot transparent (Victor, *gouvernement*), de découper un mot par syllabes en appliquant des techniques mnémoniques (Adèle, *prononciation*), de corriger la forme écrite d'un mot qui était mal orthographié (Mélissa, *traduction*), ou de donner la forme écrite d'un mot susceptible d'enrichir un CV français (Samia, *gérer*). Dans les exemples suivants, nous allons explorer les différentes façons qu'ont les apprentis-enseignants pour se focaliser sur les formes inconnues d'items lexicaux.

6.3.3.6 Donner un mot juste inconnu

Très souvent, les apprenants subissent des pannes lexicales où ils ne trouvent pas l'item lexical dont ils ont besoin pour s'exprimer. Dans ces cas, le sens est connu mais la forme leur fait défaut. Il incombe donc aux apprentis-enseignants de leur donner le mot juste. Les apprenants disposent de plusieurs moyens de manifester une panne lexicale. Un apprenant peut dire le mot en langue maternelle ou en anglais et demander une traduction, expliquer le

sens du mot recherché et demander comment cela se dit en français, rechercher une traduction en ligne et la faire valider par l'apprenti-enseignant, produire un mot qui n'est pas adapté au discours, ou donner une *formulation provisoire* (Lüdi, 1994). Dans l'exemple suivant, Catriona est en train de décrire une caricature qu'elle voit sur l'écran, mais ne semble pas connaître la forme orale du mot « employé » en français. Mélissa fait usage de plusieurs stratégies pour rendre saillant le mot juste qu'elle lui donne.

Parole orale		**Images des participantes**	**Phénomènes observés**
1 MEL 2 ALJ	[oui/] [et:] "the" travail "the" travail "the" tr- "**the employer/**" **(0,8)**		Expression d'incertitude
3 MEL 4 ALJ 5 MEL 6 ALJ	oui/ alors je vais t'envoyer le mot c'est le mot **employé** (0,5) (em[ployeur]; em[ployé])\ [tu l'as reçu dans] le: chat\ (.) c'est un employé/ oui/ (0,2) employé		Regard caméra

Exemple 28 : Mélissa utilise un commentaire méta-sémiotique

Une expression faciale d'incertitude s'observe chez l'apprenante quand elle dit « the employer » (ligne 2). Mélissa repère cette difficulté et intervient pour lui donner le mot juste, « employé », ce qu'elle qualifie de « mot » par un logonyme (ligne 3). Afin de rendre la communication plus saillante, Mélissa regarde directement vers la webcam lorsqu'elle dit « employé ». Enfin, elle fait deux commentaires méta-sémiotiques (« je vais t'envoyer le mot », ligne 3 ; « tu l'as reçu dans le chat », ligne 5). Ceux-ci aident à diriger l'attention de l'apprenante vers la fenêtre de clavardage[7] afin d'assurer le repérage de la forme écrite. Ces commentaires méta-sémiotiques s'observent souvent lors des séquences d'explication lexicale et témoignent de l'attention explicite qui est mise sur la forme des items lexicaux qui en font l'objet.

[7] Mélissa n'écrit le mot « employé » que quatorze minutes plus tard dans la séance lorsqu'elle explique la différence entre « salaire » et « salarié ».

6.3.4. Synthèse : problèmes de production

Dans cette section, nous avons analysé différentes manières de se focaliser sur la forme d'un item lexical lors d'un problème de production et les stratégies pédagogiques qui leur sont associées. La focalisation sur la forme permet aux apprentis-enseignants de :

1. Corriger la prononciation d'un mot transparent ou d'un mot dont la forme orale est partiellement connue. La focalisation sur la forme permet aux apprenants de se faire corriger par l'apprenti-enseignant et d'obtenir un enseignement personnalisé. Le dispositif permet aux apprentis-enseignants de faire usage de gestes d'informations phonologiques (Tellier, 2008a), de faire des mouvements de bouche exagérés, et de répéter par écrit certaines parties d'un item lexical ou bien l'item lexical en entier. Les apprenants bénéficient ainsi de l'input multimodal.
2. Corriger l'orthographe d'un mot qui est écrit incorrectement par l'apprenant dans la fenêtre de clavardage. La combinaison du micro et du clavardage permettent aux apprentis-enseignants d'apporter des corrections multimodales en écrivant la forme correcte dans la fenêtre de clavardage et en faisant usage de saillance vocalique pour souligner la syllabe qui a été mal orthographiée.
3. Améliorer les compétences communicationnelles des apprenants en diversifiant les situations dans lesquelles leurs connaissances lexicales peuvent être exploitées.
4. Distinguer des homophones ou mots phonétiquement proches. Ces distinctions amènent souvent à des explications susceptibles de renforcer la connaissance des mots en question. La plateforme de visioconférence facilite le dévoilement des connaissances lexicales des apprenants : ils peuvent facilement écrire et parler en même temps, ce qui rend visibles les lacunes dans leur connaissance de la forme orale et écrite des items lexicaux.
5. Rendre saillant le mot-juste qu'ils donnent suite à une panne lexicale. Les pléonasmes multimodaux, commentaires méta-sémiotiques, et mouvements de corps aident les apprentis-enseignants à le faire.

6.4 Conclusion

Au début de ce chapitre, nous nous sommes posé ces deux questions connexes : (1) comment les apprentis-enseignants se focalisent-ils sur la forme des items lexicaux qui font l'objet des séquences d'explication lexicale et (2) quelle est utilité pédagogique de cette

focalisation sur la forme ? Pour répondre à ces questions, nous avons dans un premier temps fait un survol de la littérature pour repérer quelques stratégies connues par des linguistes pour décontextualiser un item lexical. Ce faisant, nous avons compté le nombre de fois que chaque phénomène a été retrouvé dans le corpus. Dans un deuxième temps, nous avons analysé la focalisation sur la forme lors des séquences d'explication lexicale provoquées par des problèmes anticipés ou manifestés de compréhension. Dans un troisième temps, nous avons analysé la focalisation sur la forme lors des séquences d'explication lexicale provoquées par des problèmes de production. Ces analyses nous ont permis de formuler quelques réponses à ces questions.

Dans le cas de l'incompréhension d'un item lexical, en plus d'assurer une orientation mutuelle vers un item lexical qui va faire l'objet d'une explication, nous avons montré que la focalisation sur la forme permet aux apprentis-enseignants de vérifier la compréhension d'un item lexical dont l'incompréhension est anticipée, de cibler l'élément non compris dans le cas de l'incompréhension manifestée, de résoudre l'incompréhension rapidement si l'item lexical est partiellement connu, et de relier la forme au sens suite à une explication. Dans le cas d'un problème de production, la focalisation sur la forme permet aux apprentis-enseignants de corriger la prononciation et l'orthographe, améliorer les compétences communicationnelles, distinguer des homophones ou mots phonétiquement proches, et rendre saillant la forme du mot juste qui est donné suite à une panne lexicale.

Quant aux stratégies multimodales, nous avons repéré une variété de ressources qui ont été employées par les apprentis-enseignants afin de mettre en évidence la forme d'un item lexical. Nous avons retrouvé les stratégies citées par des linguistes comme les dislocations, l'autonymie, la lemmatisation, les pauses, et la saillance vocalique. Nous avons aussi repéré des stratégies multimodales comme l'acte d'écrire un mot qui fait l'objet d'une explication, l'usage de gestes pédagogiques pour donner de l'information phonologique (Tellier, 2008a), ou les expressions faciales exagérées pour rendre saillante l'émission d'un mot-clé (cf. exemple 8 en page 187). Nous avons constaté une utilisation abondante de l'utilisation du clavardage pour redoubler par écrit les mots qui font l'objet des explications, ce que nous avons appelé des pléonasmes multimodaux. Nous avons constaté également une prise de conscience de cette utilisation du clavardage, qui se manifeste par des commentaires méta-sémiotiques. Tous les apprentis-enseignants disent « je te le mets dans le chat » ou « je te l'écris », tandis que certains comme Samia et Victor ont l'habitude de dire « hop » à chaque

fois qu'ils envoient un message dans la fenêtre de clavardage. Cette focalisation sur la forme devient en quelque sorte régularisée et grammaticalisée.

Les pléonasmes multimodaux peuvent être considérés comme étant inutiles car selon le principe de redondance (Mayer, 2001; Sweller, 2005b), la répétition des mêmes informations par deux modes différents est nuisible à l'apprentissage. Or, nous sommes d'avis que dans ce contexte particulier de l'explication lexicale en langue étrangère, les pléonasmes multimodaux remplissent d'importantes fonctions pédagogiques. La forme orale et la forme écrite d'un item lexical constituent deux connaissances différentes. Pour un locuteur natif, la répétition d'un item lexical à l'oral et par écrit peut être considéré comme redondant. Or pour un apprenant qui ne maîtrise pas les formes orale et écrite de l'item lexical, l'utilisation de deux canaux livre des informations plus complètes sur la connaissance de celui-ci. Mayer et Sweller prennent d'ailleurs le soin de préciser que ce qui est redondant pour un expert ne l'est pas nécessairement pour un novice. L'utilisation du clavardage peut avoir comme effet de diriger l'attention de l'apprenant vers un item lexical et remplir ainsi la fonction méta-explicative du démarrage d'une séquence d'explication lexicale (cf. exemples 13-14 en page 192 et 193).

L'utilisation de pléonasmes multimodaux fait partie du savoir-faire multimodal. En effet, le savoir-faire multimodal consiste à savoir quels modes et quelles combinaisons utiliser et à quel moment. Par exemple, la répétition par écrit d'un mot sans commentaire méta-sémiotique peut être un bon moyen de vérifier la compréhension sans interrompre la conversation (cf. exemple 13 en page 192). La répétition par écrit accompagnée d'une pause et d'un visage d'incertitude constitue un bon moyen d'arrêter la conversation pour vérifier la compréhension. Lors d'une séquence d'explication, l'utilisation d'un commentaire méta-sémiotique est une bonne manière de diriger l'attention de l'apprenant vers une forme écrite (cf. exemple 17 en page 196). L'exemple le plus frappant de savoir-faire multimodal que nous avons analysé dans ce chapitre est l'exemple où Adèle utilise sa voix, son visage et le clavardage dans la même séquence pour se focaliser sur chaque syllabe individuellement.

Dans le chapitre suivant, nous allons explorer la façon dont les apprentis-enseignants se focalisent sur une autre facette de la connaissance d'un mot, à savoir le sens.

Chapitre 7 : Aider l'apprenant à accéder au sens d'un item lexical

7.1 Introduction

Dans le chapitre précédent, nous avons exploré la focalisation sur la forme des items lexicaux et les fonctions pédagogiques qui sont réalisées par ce type de focalisation. Le présent chapitre se donne pour vocation d'examiner la focalisation sur une autre facette de la connaissance d'un item lexical : le sens. Nous nous attachons donc à analyser la façon dont le sens des items lexicaux inconnus est expliqué multimodalement par les participants. Dans un premier temps, nous allons résumer plusieurs façons d'expliquer selon les grands noms du domaine. Puis, à l'aide d'exemples issus de notre corpus, nous allons démontrer les multiples moyens par lesquels les apprentis-enseignants expliquent le sens des items lexicaux. Nous garderons en tête tout au long de nos analyses des enjeux didactiques comme l'opportunité et l'efficacité.

7.2 Expliquer le sens d'un item lexical en langue étrangère

Plusieurs auteurs qui sont spécialistes de l'acquisition du vocabulaire ont déjà mis au jour des typologies d'explications. Nation (2013) identifie six moyens par lesquels un enseignant peut expliquer le sens d'un item lexical : réaliser des actions, montrer des objets, montrer des images, donner une définition en L1 (traduction), donner une définition en L2, donner des indices contextuels. Flowerdew (1992), quant à lui, définit trois grands types de définition : la définition formelle, la définition semi-formelle, et la substitution. Il décrit aussi une quatrième catégorie de définition : la définition ostensive. En s'inspirant de Trimble (1985), Flowerdew (1992) définit la définition formelle comme une définition qui inclut un terme, sa classe d'objets et une caractéristique qui la distingue[1]. La définition semi-formelle est identique à la définition formelle mais n'inclut pas la classe d'objets. Parfois la classe n'est pas nécessaire car elle est vague ou redondante. La substitution a lieu quand un mot, partie d'un mot, ou phrase(s) ayant un sens similaire est substitué pour le nouveau terme. Flowerdew définit trois types de substitution : la synonymie, la paraphrase et la dérivation. Un *synonyme* est un item lexical ayant un sens similaire ; une *paraphrase* est une phrase ou des

[1] Par exemple, « Un thermomètre (terme) est un instrument (classe d'objets) qui est utilisé pour mesurer la température (caractéristique qui le définit) » (Flowerdew, 1992).

phrases avec un sens similaire ; une *dérivation* est quand un mot ou partie d'un mot, souvent du latin ou du grec, est cité pour expliquer la dérivation et donc le sens du terme auquel il fait référence. Ces trois types de substitution peuvent être combinés avec une définition formelle ou semi-formelle. La définition ostensive, qui est minoritaire selon Flowerdew, consiste à indiquer un stimulus visuel comme un objet, une photographe ou un diagramme. Le geste n'est pas mentionné. Cette catégorisation a des points commun avec celle de Fasel Lauzon (2014) qui a recensé quatre types d'explication : la traduction, la synonymie, la définition, et la mise en situation. La mise en situation, selon elle, se caractérise par une définition qui commence par « c'est quand … ». Fasel Lauzon, contrairement à Flowerdew et à Nation, ne mentionne pas la multimodalité, parlant uniquement de définitions linguistiques.

Du court survol de la littérature ci-dessus, nous retenons trois grandes catégories d'explication : l'explication en français, l'explication par traduction et/ou recours à la langue maternelle des apprenants, et l'explication ostensive/visuelle. Si les grandes catégories d'explication ont déjà été dégagées, il reste à savoir, comme dans le chapitre précédent, comment les apprentis-enseignants s'emparent des affordances du dispositif Visu afin d'expliquer les sens des items lexicaux inconnus. Nous allons donc diviser ce chapitre en deux grandes sections : l'explication voco-verbale et l'explication ostensive. La première section sera divisée encore entre explications en français et explications en L1. Dans la seconde section, nous analyserons la façon dont les apprentis-enseignants donnent des explications ostensives à l'aide de leurs corps et des outils du dispositif.

Cette division peut sembler contradictoire à notre plaidoyer dans le chapitre 3 pour une vue holistique des ressources sémiotiques. Bien entendu, ces deux catégories ne sont pas mutuellement exclusives : nous opérons cette distinction uniquement pour faciliter la présentation des analyses. Plusieurs stratégies d'explications par plusieurs modes différents peuvent coexister et nous verrons que c'est souvent le cas.

7.3 Explications voco-verbales

Nous avons repéré plusieurs types d'explications voco-verbales, et nous les avons divisés en deux groupes : le recours à la langue maternelle des apprenants et les explications en français.

7.3.1 Le recours à la langue maternelle

Le recours à la langue maternelle de l'apprenant est très efficace pour expliquer le sens d'un item lexical (cf. chapitre 2, section 2.4.2.5 en page 56). Dans notre corpus, l'utilisation d'anglais abonde dans toutes les phases des explications aussi bien par les apprenants que par les enseignants. La langue maternelle de Catriona (apprenante de Mélissa) est l'espagnol, et elle utilise parfois cette langue pour communiquer avec Ana, son binôme.

7.3.1.1 Usage de l'anglais par les apprenants

Les apprenants utilisent l'anglais beaucoup plus que les apprentis-enseignants, surtout Ana et Alannah. Le recours à la langue maternelle se fait très peu par écrit. On note que le nombre de séquences dans lesquelles il y a recours à la langue maternelle par les apprenants est plus de deux fois plus que le nombre de séquences dans lesquelles les apprentis-enseignants recourent à l'anglais (cf. Figure 2). Puisque le sujet de notre thèse porte sur les stratégies employées par les apprentis-enseignants, nous n'allons pas détailler à l'aide d'exemples l'utilisation d'anglais par les apprenants. Nous pouvons tout de même résumer les fonctions remplies par le recours à la langue maternelle par les apprenants et les difficultés que cela peut provoquer.

Apprenti-enseignant	Adèle		Mélissa		Victor		Samia		Emilie		Séverine		Étienne		Total
Traduction orale (apprenant)	Alannah	14	Ana	18	Liam	8	Angela	5	Aiden	4	Naomi	5	Conor	1	69
	Catriona	0	Alejandra	8	Conor	2	Sean	0	Fiona	5					
	Total	14	Total	25[2]	Total	10	Total	5	Total	9					
Traduction écrite (apprenant)			Ana	2											3
			Alejandra	1											
			Total	3											
Total traductions apprenants	14		28		10		5		9		5		1		72

Figure 1 : Le recours à la langue maternelle par les apprenants

Le recours à l'anglais peut parfois être utile quand les apprenants utilisent leur langue maternelle pour signaler implicitement ou explicitement des pannes lexicales, pour se faire aider, et pour vérifier leur propre compréhension d'un item lexical à la suite d'une séquence d'explication. Cependant, les apprentis-enseignants risquent de ne pas comprendre la langue

[2] Dans une des séquences, Ana et Alejandra font toutes les deux une traduction. Nous n'avons pas compté cette même séquence deux fois.

maternelle des apprenants, ou de ne pas comprendre que c'est un appel à l'aide. En effet, un apprenti-enseignant peut accepter une phrase en anglais sans donner la formulation en français. Un appel à l'aide peut être plus perceptible s'il se fait par écrit. A un moment, une apprentie-enseignante croit à tort qu'elle comprend un mot en anglais et elle donne une traduction erronée. Enfin, l'incompréhension d'un mot anglais par un apprenti-enseignant peut avoir de bonnes conséquences didactiques si l'apprenti-enseignant demande à l'apprenant d'expliquer en français l'item lexical recherché.

7.3.1.2 La traduction comme stratégie d'explication

L'utilisation de la langue maternelle des apprenants par les apprentis-enseignants reste une stratégie d'explication observable qu'il convient d'analyser. L'utilisation de traductions par les apprentis-enseignants se résume comme suit :

	Adèle	Mélissa	Victor	Samia	Emilie	Séverine	Étienne	Total
Traduction orale	6		5	1	5	5		22
Traduction écrite	1				6			7
Traduction indirecte (orale)					2			2
Total traductions apprentis-enseignants	7	0	5	1	13	5	0	31

Figure 2 : Utilisation des traductions par les apprentis-enseignants

Tous sauf Mélissa et Étienne utilisent des traductions pour expliquer, et la plupart des traductions se font à l'oral. Séverine utilise la traduction plus fréquemment que les autres apprentis-enseignants à raison de cinq traductions par séance. Nous avons inclus la catégorie « traduction indirecte » pour Emilie qui réfère à l'anglais sans dire la traduction. Par exemple, en parlant du mot « employé » lors de la première séance, elle dit : « c'est un mot transparent je crois qu'en anglais vous avez la même chose ». En parlant de l'expression « tenir les délais » lors de la quatrième séance, elle dit « alors euh les délais c'est pas exactement comme en anglais même s'il y a le rapport au temps aussi » (cf. exemple 7 en page 225). Ainsi, elle active la L1 des apprenants sans prononcer les traductions. Dans les sous-parties qui suivent, nous allons analyser les fonctions didactiques des traductions employées par les apprentis-enseignants.

7.3.1.2.1 La traduction pour résoudre rapidement l'incompréhension

Les apprenants emploient des traductions assez souvent, et leur usage premier consiste à expliquer le sens d'un mot inconnu. Lors de la deuxième semaine, Adèle donne des adjectifs à ses apprenantes qu'elle juge utiles pour un CV de gardien d'enfants, dont le mot « efficace » qui provoque une demande d'aide de la part de Catriona :

Transcription de la parole		
1	ADL	efficace/
		<((clavardage)) efficace>
		(0,7)
2	CTR	efficace c'est quoi (0,3) (°pardon°)
		(0,9)
3	ADL	euh efficace/ c'est "efficient"\
4	CTR	ah oh ok\
		(1,5)
5	ADL	quand tu dis euh: "I'm efficient"\ (.) je suis effic[ace/]
6	CTR	[ok] (.)
		efficace

Exemple 1 : La traduction seule

Une traduction semble adaptée comme stratégie où la compréhension rapide et le volume de mots transmis sont de première importance. Ici, Adèle cherche à agrandir le nombre de mots que ses apprenantes pourraient mettre sur leurs CV. Puisque les apprenantes connaissent déjà le sens de ces mots dans leur langue maternelle, une traduction est adaptée.

La traduction peut aussi être combinée avec d'autres stratégies, comme la définition. La traduction servirait de renforcement dans le cas où la définition serait comprise. Dans l'exemple suivant, Victor donne une définition et ajoute une traduction du mot « pourboire » pour expliquer, assurer la compréhension, partager des connaissances culturelles et renforcer la mémorisation. L'explication émerge d'une conversation sur les expériences professionnelles de Conor qui travaille dans un restaurant.

Transcription de la parole		
1	VIC	vous avez des pourboires/ (2,1) tu sais c`que [c'est pourboi]re/
2	CON	[des:]
		(1,5)
3	VIC	<((clavardage)) pourboire>
4	CON	°pourboires:° (1,9) non
		(0,6)
5	VIC	c'est hum: (.) j`vais t'expliquer d'abord en français/ puis
		ensuite j`te donnerai le le mot en anglais/ (0,7) c'[est hu]m:&
6	CON	[oui]
7	VIC	&la somme d'argent/ que les gens laissent/ en surplus de: la
		fac[ture hum]
8	CON	[a- ah:]
9	VIC	donc le "ti[p]"
10	CON	[(alors)] (.) oui (0,3) "tip" oui [c'est] (1,2) hum:&
11	VIC	[oui]
12	CON	&nous nous avons euh (0,3) un système que: (1,1) nous (.) doive

```
          (.) euh (.) partager/
          (0,8)
13 VIC    d'accord\
          (0,2)
14 CON    nos nos pourboires avec (.) toute (.) euh: (0,8) toute le "staff"
15 VIC    d'accord\
16 CON    "so" c'est hum euh: (0,4) c'est un domma- un dommage mais (0,5)
          c'est le système "so" [c-]
17 VIC                          [hm] hm hm (0,3) oui d'accord
          (0,4)
18 CON    [(°euh c'est°)]
19 VIC    [ici] en france/ hum: dans les restaurants le pourboire/ (0,6) il
          est inclus dans la facture en fait\
          (1,3)
20 CON    ah oui (0,4) [c'est c'est le même ch]ose euh aux états-unis
21 VIC                 [donc euh:]
          (0,3)
          ah d'accord je savais pas\ (0,4) d'accord d'accord
          (0,5)
22 CON    °hm hm°
23 VIC    et c'est c- je trouve que c'est un peu dommage parce que: (0,4)
          le serveur finalement euh peu importe euh: (0,7) s'il sourit/
          s'il est sympathique/ euh il obtient son pourboire donc euh\
24 CON    hm oui
```

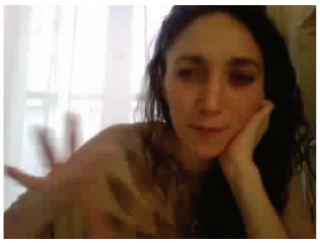
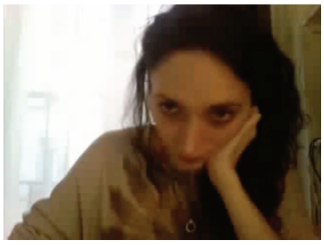
Exemple 2 : La traduction combinée avec d'autres stratégies d'explication

Victor, bien conscient de son usage de stratégies d'explication (ligne 5), décide de donner une définition formelle (Flowerdew, 1992) suivie d'une traduction. Il annonce son usage de stratégies, ce qui pourrait avoir des effets contre-productifs. Notamment, Conor pourrait décider qu'il n'a pas besoin de l'explication en français car il sait que la définition va suivre. Cependant, l'explication en français semble avoir fonctionné, car Conor réagit en donnant un marqueur de compréhension. Il dit « ah », s'éloigne de l'écran et hoche la tête (ligne 8). Puis, il ratifie la traduction donnée par Victor (ligne 10).

Ensuite, après avoir entendu la traduction, Conor parle de son expérience personnelle avec les pourboires, en parlant de la politique de son restaurant (lignes 12-14) et y porte un jugement (ligne 16). Victor encourage cette co-construction du sens en donnant des régulateurs (lignes 13, 15, 17) et des hochements de tête. Puis, il partage des connaissances culturelles sur la coutume française concernant le pourboire (ligne 19), valide l'intervention de l'apprenant quand il ajoute des connaissances culturelles (ligne 21) et porte un jugement sur la politique française concernant le pourboire (ligne 23). Porter jugement sur un mot semblerait favoriser la mémorisation (Bourdon-Larger, 2008). Il est possible que la traduction « tip » ait activé des connaissances culturelles de l'apprenant et l'ait poussé à y porter un jugement.

Une dernière fonction de la traduction que nous avons repérée chez les enseignants est la traduction pour vérifier la compréhension de l'apprenant. Ceci peut se faire suite à une

explication ou pendant une tâche communicative où l'enseignant souhaite s'assurer que l'apprenant comprenne. Dans l'exemple suivant, Adèle parle de ses habitudes à Lyon quand elle utilise une traduction pour vérifier la compréhension de ses apprenantes :

Transcription de la parole	Image du participant	Phénomènes observés
1 ADL et hum: pour la charcuterie/ je vais aussi/ dans une boucherie/ (0,8) où hum: ils reçoivent/ (.) euh de la charcuterie de la région de lyon (1,9) et le fromage aussi\ des petites **fermes (0,2) "farmer"/ (0,2) ferme/ (1,0)**		Expression d'incertitude et geste métaphorique : main ouverte vers le bas, faisant de grands mouvements circulaires
2 ALN "yeah" (0,3) ferme\ **(0,9)** 3 ADL donc tout est sur lyon\		Hochement de tête

Exemple 3 : La traduction pour s'assurer de la compréhension des apprenants

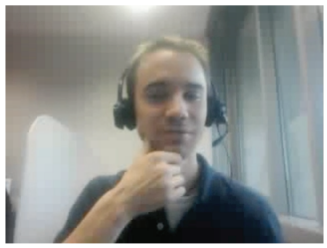
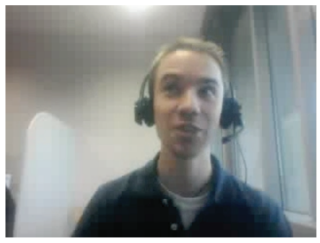
En traduisant le mot « ferme », Adèle s'assure que ses apprenantes la suivent pendant qu'elle décrit ses habitudes à Lyon. L'alternance codique est aussi un moyen de souligner un item lexical (Miecznikowski, 2005). Il est intéressant de noter qu'Adèle décide de traduire le mot « ferme », qui est un mot à très haute fréquence, alors qu'elle ne souligne pas le mot « charcuterie », qui est employé moins souvent.

Les traductions servent donc à expliquer les sens des mots inconnus ou à vérifier la compréhension de l'apprenant. Cette stratégie se révèle utile quand une compréhension rapide est nécessaire, quand l'apprenti-enseignant estime que c'est un mot dont les apprenants connaissent le sens dans leur langue maternelle, et quand l'apprenti-enseignant connaît la traduction. Une traduction peut aussi servir à activer des connaissances culturelles et à provoquer davantage d'output (exemple 2). Nous verrons dans le chapitre 9 lorsqu'Adèle explique le mot « bobo » que parfois les apprentis-enseignants donnent des traductions erronées.

7.3.1.3 Attitudes divergentes des apprentis-enseignants à propos du recours à la langue maternelle des apprenants

Par des commentaires méta-pédagogiques et méta-explicatifs, les enseignants révèlent leurs attitudes à l'égard de la traduction. Victor, par exemple, croit que c'est une stratégie

d'explication utile, comme en témoignent ses commentaires au cours de l'explication suivante. La deuxième activité de la première séance s'intitule « Les Français et le droit de grève » et les apprentis-enseignants avaient prévu de demander aux apprenants si les Irlandais faisaient aussi la grève. L'explication lexicale qui suit est composée de deux parties qui sont séparées de 3 minutes et 24 secondes. Pendant ce temps, deux autres séquences d'explication lexicale surgissent. Ce n'est que plus tard que Victor se rend compte que la première partie n'a pas été efficace. Ces deux parties illustrent d'une part l'importance que l'apprenti-enseignant accorde à la traduction et d'autre part l'efficacité de cette stratégie d'explication.

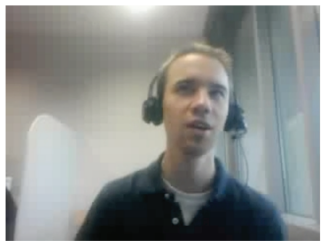
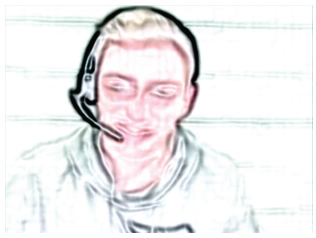
Transcription de la parole			**Image du participant**	**Phénomènes observés**
1	VIC	est-ce qu'ils font euh souvent la grève/ ou euh\ (0,2) pas trop\ pas autant qu'en france sans doute\ (1,2)		Victor touche son menton lorsqu'il vérifie la compréhension
2	LIM	euh: pardon/ (0,7)		
3	VIC	ils font pas souvent la grève\ (.) j'imagine\ (1,8) tu sais ce que **[c'est la grè]ve/**		
4	LIM	**[euh:] (1,3) euh non (0,4)**		
5	VIC	**ah d'accord\ (0,4)**		
		c'est q[uand les ge]ns/ (.)&		Geste iconique pour illustrer l'acte de scander des slogans
6	LIM	[°euh°/]		
7	VIC	&arrêtent de de travailler/ (.) et ils sortent **dans la rue/ pour scander des (.) des** slogans:/ et\ (1,0) (et) vouloir de: de meilleurs conditions de travail\ (0,6) faire la grè[ve]\		
8	LIM	[o]k- (1,2)		
9	VIC	en anglais/ **je ne sais pas ce que c'est donc euh** (1,7) mais tu: tu comprends\ (1,1)		Haussement de sourcils et regard pensif
10	LIM	oui e[uh je pe]nse/ que\		
11	VIC	[d'ac-] (2,6)		
12	LIM	hum (5,3)		
13	VIC	oh euh: d'accord\ donc t'as répondu/ ça va/		

Exemple 4 : Un manque de traduction commenté

A la ligne 2, Liam signale son incompréhension mais ne dit pas quel élément il n'a pas compris, et par conséquent Victor reformule sa question sans s'attarder sur le mot « grève ». Suite à une pause de 1800 millisecondes, Victor vérifie la compréhension du mot « grève » et touche son menton en attendant une réponse. Quand l'item lexical non compris est identifié et

l'orientation mutuelle sur celui-ci est établie, Victor, qui utilise un vocabulaire peu fréquent, donne une définition lexicographique accompagnée d'un geste iconique pour illustrer l'acte de scander des slogans (lignes 5-7). Ce qu'il dit à la ligne 9 est révélateur de ses croyances pédagogiques : il pense qu'une traduction serait un bon moyen d'expliquer l'item non compris. Ce commentaire est d'autant plus intéressant qu'il vient après sa définition qui est censée avoir résolu l'incompréhension. Victor vérifie la compréhension à la fin et accepte la réponse tiède de son apprenant, bien que « L'apprenant, de son côté, [doive] fournir à l'enseignant la preuve qu'il a intégré de nouvelles connaissances » (Cicurel, 2011). On verra plus loin que la compréhension de Liam n'était pas du tout suffisante. On peut donc formuler la proposition didactique suivante : ne pas accepter une réponse aussi incertaine à une vérification de compréhension ; exiger que l'apprenant fasse montre de sa compréhension en reformulant, en expliquant, en donnant une réponse appropriée à la tâche (Smith, 2003), voire en traduisant.

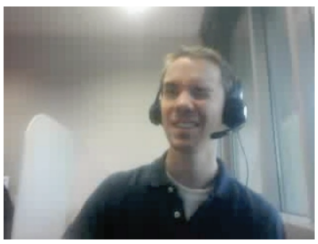
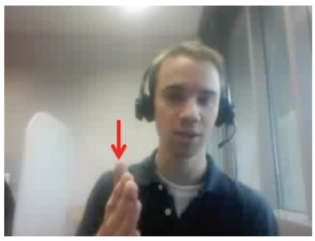
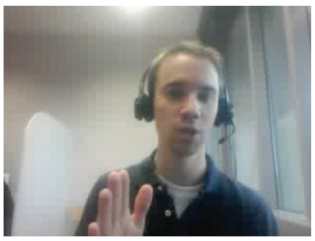
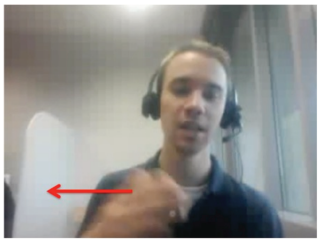
Avant la deuxième partie de l'explication du mot « grève », Victor vérifie la compréhension du mot « manifestation ».

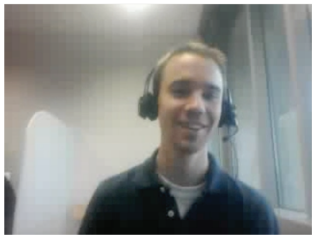
Transcription de la parole			**Image du participant**	**Phénomènes observés**
1	VIC	hum: hop: (1,9) maintenant/ c'est (.) les français/ et le droit de grève\ (1,6) euh: donc je t'ai demandé toute à l'heure si tu connaissais le mot grève/ (.) tu m'as répondu que non/ mais que: (0,5) donc euh maintenant oui/ je t'ai expliqué euh: (.) ce que c'était/ (1,0) est-ce que tu fais **une différence/ (.) entre le mot grève/ (.) et le mot/** (.) manifestation\ (2,7)		Regard pensif
2	LIM	hu:m (2,1)		Sourire de satisfaction
3	VIC	est-ce que tu sais ce que c'est une manifestation/ (0,9)		
4	LIM	euh oui je pense/ (.) euh (1,0) euh "protest"/ (0,5)		
5	VIC	mh mh/ (.) oui oui (.) c'est ça\		
6	LIM	**°oui (.) oui°/**		

Exemple 5 : Accepter une définition en anglais de l'apprenant

Cet exemple est révélateur pour plusieurs raisons. Victor pense avoir réussi à faire comprendre le sens du mot « grève » à son apprenant (ligne 1). Ce n'est pas le cas, ce qui

nous a incité à faire la proposition didactique précédente. La capacité à repérer et à anticiper l'incompréhension chez l'apprenant semble donc être une compétence à développer chez cet apprenti-enseignant. Deuxièmement, on voit que Victor accepte la traduction anglaise donnée par Liam comme preuve qu'il comprend l'item lexical en question, comme on l'a vu précédemment (exemple 2). Tout de suite (940 ms) après cette vérification de compréhension, Liam montre à nouveau son incompréhension du mot « grève ».

Transcription de la parole		**Image du participant**	**Phénomènes observés**
1	LIM	hum: (1,7) euh qu'est-ce que c'est: "or" (0,4) qu'est-ce que c'est euh (.) le grève/ **(1,5)**		Rapprochement de l'écran et regard d'incertitude
2	VIC	alors (.) la grève hum: (1,3) hum: (0,7) j'ai oublié le sens en anglais en fait c'est hum: **.tsk (2,3)** ah: tu sais c'est quand les gens euh: par exemple: (0,6) chais pas euh hum: (1,7) les gens d'un hôtel/ les les les employés		Yeux fermés, un sourcil levé
		d'un hôtél/ (.) décident (.) d'arrêter (0,5) de: travailler/ (.) un jour do**nné/ (0,2) ils s'entendent/ (0,2)**		Geste emblématique, mouvement vers le bas puis tenu
		ce jour là on travaille pas/		Geste emblématique tenu
		(0,7) et ils font grève (1,5)		Battement vers la gauche de l'écran

		ah oui je sais en anglais c'est euh: (0,3) "strike (0,9)			Sourires
3	LIM	AH:\ (.) oui\ (0,3)			
4	VIC	voilà\ (0,5) et [donc la différe- (0,3) la]			
5	LIM	[je comprends]			

Exemple 6 : Une traduction est plus efficace qu'une explication en français

Pour sa deuxième explication, Victor donne un exemple au lieu d'une définition lexicographique. Cette explication, parsemée de gestes, ne réussit toujours pas à résoudre l'incompréhension. Suite à une pause de 1,5 secondes, Victor se rappel enfin de la traduction, ce qui résout instantanément l'incompréhension. Cet exemple est parlant car il illustre un cas où la combinaison d'une définition et d'un exemple, tous les deux avec gestes, est manifestement moins efficace qu'une simple traduction. Il illustre aussi l'importance que Victor accorde à cette stratégie.

Ultérieurement durant le semestre, lors d'un débriefing, Victor s'exprime à ce sujet, en disant que :

> Mais, à la toute première séance, quand je… j'ai essayé de lui expliquer ce que c'était « la grève », j'ai expliqué deux, trois fois. Après, j'ai dit « ben, strike ». Son visage s'est allumé… Donc, parfois, ça fonctionne, parfois… c'était nécessaire… parfois !

On peut se demander donc quelle sorte de contrat didactique Victor met en place avec son apprenant. Pour lui, comme pour Séverine et Adèle, la traduction est autorisée, voire recommandée. Pour d'autres apprentis-enseignants, qui sont moins à l'aise en anglais comme Mélissa, la traduction est « en dernier recours » comme dit Mélissa lors du débriefing.

Cette attitude positive à l'égard de la traduction n'est pas partagée par tous les apprentis-enseignants. Emilie, par exemple, exprime de la culpabilité quand elle traduit l'expression « tenir les délais » après l'avoir définie. Cette explication a lieu lors de la quatrième semaine quand les participants font état des compétences requises pour la planification d'un goûter d'anniversaire.

Transcription de la parole		
1	EMI	et surtout/ (0,5) une expression assez importante/ c'est tenir les délais\ (.) est-ce que vous la comprenez celle-ci/ (2,0)
2	SOH	euh "no" qu'est-ce que c'est les (.) délais (0,6)
3	EMI	alors (.) euh les délais c'est pas exactement comme en- comme en anglais/ (.) même s'il y a le rapport au temps aussi/ (0,6) tenir un délais/ (0,4) ça veut dire/ (.) euh être à temps\ (0,2) vous devez faire les choses/ (.) dans les temps\ (.) c'est à dire/ (.)

		au moment dit\ (0,3) vous n`pouvez pas avoir de retard\ (0,8)
4	SOH	oh oui (0,2)
5	EMI	donc si on vous demande (.) de tenir un délais/ (0,3) vous devez être (0,4) à l'heure\ (.) vous devez le faire au bon moment\ (1,6) hop/ (3,5) <((clavardage)) tenir un délai = be just in time> (0,9) voilà\ j`vous l`mets dans l`tch[at je s]uis p[as/ cen]sée&
6	SOH	[oh oui] (0,3) [oui]
7	EMI	&vous l`faire mais voilà j`pense que ça s`ra p`t-être plus claire\ (1,4) vous allez faire les [choses à temps\] (.)&
8	SOH	[oui je comprends]
9	EMI	&d'accord/ (0,5)
10	SOH	oui

Exemple 7 : Emilie se culpabilise suite à une traduction

Emilie donne une définition en français de l'expression (lignes 3 et 5). Puis, elle la traduit par écrit (ligne 5). Ce qui est intéressant ici est que son attitude à l'égard de l'usage de traductions est révélée quand elle dit qu'elle n'est pas censée le faire (lignes 5-7). Pour elle, le recours à la traduction est répréhensible. En effet, elle utilise la traduction à l'écrit comme si c'était quelque chose de clandestin et interdit. Cet épisode de *mode switching* (Sindoni, 2013) pour l'aparté métalinguistique met plus de distance entre son explication orale et la traduction écrite.

Cette opinion est partagée par Mélissa, qui ne veut pas qu'Ana, son apprenante la plus forte, aide Alejandra en espagnol. Lors de la deuxième semaine, quand Alejandra (d'origine espagnole) rencontre un problème de compréhension, Ana (d'origine méxicaine et canadienne), intervient ainsi :

		Transcription de la parole
1	ANA	hum (.) je peux dire (.) quelque chose/ (0,3)
2	MEL	oui/
3	ANA	(0,4) ouais hum alejandra si tu n'entends pas euh quelques mots/ ou quelque chose en anglais/ ou en français/ (.) tu me peux: (0,5) euh je peux te traduire en espagnol parce que je sais que il y a beaucoup de mots que: (0,6) souvent (.) tu n'entends pas alors euh si tu me demandes en espagnol/ "si me preguntas en espagnol yo te lo digo y te traduzco"\ (0,5)
4	ALJ	<((en riant)) o[k ((rire))] (0,3) ((rire))
5	MEL	[((rire))] mais (0,6)
6	MEL	c'est mieux si si (0,3) si tu peux le: l'expliquer en français ana\ (0,6)
7	ANA	ouais
8	ALJ	°((rire))°
9	MEL	si tu peux expliquer le mot en français c'est mieux

Exemple 8 : Mélissa n'apprécie pas l'usage de traductions en espagnol

Ana offre de l'aide en espagnol (ligne 3), mais Mélissa répond (lignes 6, 9) que l'usage de français est préférable pour l'explication. Fasel Lauzon (2014) a trouvé aussi que les enseignants dans son étude pensaient que la L1 ne devrait pas être utilisée en classe de langue. Ana tente de redéfinir le contrat didactique de façon à ce que l'espagnol soit acceptable entre les apprenantes, mais Mélissa rejette cette proposition. Mélissa redéfinit le contrat didactique en demandant que les explications se fassent uniquement en français.

Dans cette section, nous avons vu que les traductions sont utilisées par les enseignants comme la seule stratégie d'explication, ou en combinaison avec d'autres. Dans certains cas, la traduction est la seule stratégie qui parvient à résoudre l'incompréhension d'un item lexical. Les attitudes des apprentis-enseignants envers l'usage de cette stratégie sont mixtes. Alors que la traduction est une stratégie privilégiée par Victor, Adèle, Emilie et Séverine, Samia, Mélissa et Etienne ne sont pas du même avis. Samia se culpabilise quand elle donne une traduction, et le recours à d'autres langues met Mélissa mal à l'aise. Une variété de contrats didactiques et d'attitudes s'observe.

7.3.2 Explications en français

La plupart des explications repérées dans notre corpus se font en français. Comme le rappelle Bogaards (1994), l'usage de définitions, paraphrases ou autres types de définitions en L2 est avantageux car il renforce des liens entre éléments de la L2 et permet ainsi à l'apprenant de se construire un réseau qui est indépendant de celui de la L1 et d'utiliser la L2 de façon communicative, ce qui est motivant. Un des inconvénients de l'explication en L2 est que le risque d'incompréhension est plus élevé, surtout en début d'apprentissage. Si on utilise des paraphrases, il est difficile de ne pas employer des mots de basse fréquence, et « comme les paraphrases ne sont pas supposées constituer de vraies définitions, elles fournissent en général des informations tronquées et demandent à être complétées par d'autres sources » (Bogaards, 1994: 169). Ce dernier point n'est pas problématique *per se*, car il est impossible d'apprendre un mot en une seule rencontre : on n'apprend que des morceaux de la connaissance d'un item lexical à la fois, et le reste se complète lors de rencontres ultérieures (Nation, 2013).

Nous avons compté le nombre de séquences d'explication lexicale dans lesquelles se trouvent des définitions générales, définitions liées à la tâche, synonymes, exemples et reformulations. L'utilisation de ces stratégies d'explication en français se résume comme suit :

	Adèle	Mélissa	Victor	Samia	Emilie	Étienne	Séverine	Total
Définition générale orale	8	15	17	10	12	2	8	72
Définition orale (tâche)	1			2				3
Synonyme oral	4	7	3	4	8	1	2	29
Exemple oral	6	6	5	2	5	1	2	27
Reformulation orale	2	2		1				5
Définition écrite		1		1			1	3
Synonyme écrite		1			3		1	5
Total	21	32	25	20	28	4	14	144

Figure 3 : Les explications en français

Les analyses suivantes suivent la classification ci-dessus.

7.3.2.1 Définitions générales

La plupart de la focalisation sur le sens se fait par des définitions générales, c'est-à-dire des définitions qui sont décontextualisées et qui peuvent être comprises hors contexte de la tâche en cours.

7.3.2.1.1 Vérifier la compréhension d'un item lexical par l'apprenant

Les définitions courtes et auto-déclenchées sont souvent employées par les apprentis-enseignants lorsqu'ils expliquent la consigne d'une tâche afin qu'elle soit comprise.

Transcription de la parole		**Image du participant**	**Phénomènes observés**
1	MEL en fait/ j`vais jouer le rôle/ euh (0,4) d'un: d'une **restauratrice/** (.) c'est-à-dire quelqu'un qui: fait la cuisine/ (.) donc cuisinière/ ((17 secondes plus tard)) 2 MEL donc\ (.) je vous envoie les mots-clés\ (0,6) je suis restauratrice/ (.) c'est-à-dire cuisinière/		Regard caméra et geste déictique vers soi

Exemple 9 : Une définition simple et rapide

La priorité principale pour Mélissa dans cet exemple est que la consigne de l'activité soit comprise. Ses regards directs vers la caméra témoignent de l'attention qu'elle porte à l'efficacité de la communication. « Restauratrice » étant un des mots-clés de l'activité, elle le définit deux fois sans que les apprenants ne le lui demandent à l'aide d'un geste. La combinaison de sa voix et son geste renforce la compréhension. Un autre exemple d'une explication simple, rapide et liée à la consigne de l'activité a été présenté dans le chapitre précédent où Étienne explique « paresseux » (exemple 18 en page 197). Les explications courtes, simples et auto-déclenchées lors des consignes assurent que l'activité soit comprise.

7.3.2.1.2 Reprendre une définition de l'apprenant pour donner un mot juste

Quand un mot juste est donné par l'apprenti-enseignant, l'apprenant doit l'intégrer dans son lexique mental. L'apprenti-enseignant peut faciliter cette intégration en partant de ce que l'apprenant sait déjà. Dans l'exemple suivant, Samia répète la définition donnée par l'apprenante.

Transcription de la parole	
1 ANG	j'ai un parapluie: et:
2 SAM	((rires))
3 ANG	une euh: (1,3) ouais\ (2,9) c'est vraiment important d'avoir une: (0,4) veste euh: pour la plu- (.) pluie i[ci/ ((rire))]
4 SAM	[<<((en riant)) d'accord>>] (0,5) ((rire)) donc un imperméable/ (0,7) ça s'appelle un IMPERMÉABLE
5 ANG	ah:
7 SAM	<((clavardage)) imperméable> (3,6) en français\ (.) une veste pour la pluie\
9 ANG	ouais

Exemple 10 : Répéter la définition de l'apprenant

Le mot « parapluie » fait défaut à Angela, qui met en place la *formulation provisoire* (Lüdi, 1994) « veste pour la pluie ». Samia remplit son rôle que le contrat didactique lui attribue et comprend à la ligne 4 qu'il s'agit d'un appel à l'aide pour trouver le mot juste. Samia le lui donne avec saillance vocalique (ligne 4), les logonymes « appeler » (ligne 4) et « français » (ligne 7) et renforcement par clavardage (ligne 7). Samia va plus loin et renforce le lien entre le mot juste et la formulation provisoire en répétant la définition de l'apprenante. Le discours rapporté peut d'ailleurs être accompagnée de guillemets gestuels (cf. chapitre suivant).

7.3.2.2 Définitions contextualisées par rapport à la tâche

En expliquant le sens d'un item lexical, les apprentis-enseignants ont des décisions à prendre quant au degré de généralité de l'explication. Un apprenti-enseignant peut décider de ne pas sortir de la tâche en expliquant, soit l'exemple suivant :

Transcription de la parole		
1	SAM	<((clavardage)) 6) Mettre en place un budget pour payer les frais du camion et de larestauration et le salaire des employés> (10,9)
2	SEN	qu'est-ce que c'est les: (.) {frɛz} de camion (1,2)
3	SAM	les frais/ (0,6) les frais [c'est] en fait l'argent euh: dont&
4	SEN	[oui]
5	SAM	&j'aurais besoin pour pa- payer l[e:]
6	SEN	[ah d'accord\] (0,4)
7	SAM	le: [le le:]
8	SEN	[pour la coût de:]
9	SAM	<((clavardage)) 7) Budget pour la campagne marketing> (0,7)
10	SEN	nourr[iture/]
11	SAM	[voi]là\ (0,5) l'essence/ la nourriture/ euh:&
12	SEN	°oui oui oui°
13	SAM	&l'essence du camion/ par exemple/ (0,7) e[uh] la nourriture&
14	SEN	[oui]
15	SAM	&euh: pour payer par exemple si j'ai pas de camion que j'ai pas d'argent peut-être que je vais devoir LOUER un camion/ (0,3)
16	SEN	mh\ (0,6)
17	SAM	donc tout ça en fait\

Exemple 11 : Une définition liée à la tâche

Samia, alors en train d'écrire les étapes à suivre afin de mettre en place un projet de camion-restaurant, écrit « payer les frais », ce qui déclenche une séquence d'explication lexicale. Sean ne demande pas une explication de l'item lexical « frais », mais une explication des « frais de camion ». Samia décide alors de donner une explication liée à la tâche, en donnant des exemples de frais dans le cadre du projet. Sean semble avoir compris, car il contribue à la séquence de négociation du sens en ajoutant « pour le coût de nourriture » (lignes 8, 10), proposition qui est reprise par Samia (ligne 11).

Cette stratégie a plusieurs avantages. Hormis la rapidité d'une explication qui est faite uniquement pour l'activité en cours, une explication qui fait intervenir à la fois la forme, le sens et la relation entretenue avec l'activité amène à un traitement approfondi de l'item lexical (Bogaards, 1994). Cet auteur (1994 : 92) explique que :

> « C'est la tâche à exécuter qui impose le niveau de traitement : ainsi, la simple répétition d'un élément n'implique que des niveaux superficiels, tandis que les tâches difficiles comme celles qui font appel à des facultés de raisonnement ou à la comparaison détaillée, et qui font donc intervenir le sens du matériel utilisé, ont lieu à un niveau profond ».

Grâce à cette explication, Sean a accès à la fois à la forme écrite (ligne 1), la forme orale (ligne 3), et à son usage réel par rapport à la tâche (lignes 11-15).

Un autre avantage des explications fortement liées à la tâche est que lors des rencontres ultérieures, l'apprenant sera obligé d'utiliser l'item lexical de manière générative

(Joe, 1995, 1998) (cf. chapitre 2, section 2.4.3.3 en page 61). C'est-à-dire l'utiliser dans un contexte autre que celui dans lequel il a été appris, ou utiliser une autre acception de l'item lexical que celle qui a été initialement apprise. Ce type d'usage est propice à l'acquisition du vocabulaire et constitue une partie de la compétence lexicale.

7.3.2.3 Synonymes

Les synonymes oraux sont présents dans vingt-neuf séquences d'explication lexicale, et les synonymes écrits sont présents dans quatre séquences. La synonymie est la deuxième stratégie la plus utilisée après la définition. Il est difficile d'employer des synonymes parce que de vrais synonymes sont rares, et que les synonymes fournis en langue étrangère ont moins de représentations chez les apprenants que les traductions en langue maternelle (Bogaards, 1994). Les synonymes ne marchent pas dans tous les contextes, n'ont pas le même comportement syntaxique et n'appartiennent pas toujours au même registre de langue (Tréville & Duquette, 1996). De plus, la synonymie n'est pas adaptée à assurer une compréhension profonde des items lexicaux. Mortureux (2008 : 81) explique que :

> « La signification d'un mot n'est pas un bloc homogène, la meilleure façon de définir un mot ne peut être de lui trouver un synonyme, ou du moins un équivalent, mais plutôt d'énumérer dans une paraphrase les sèmes que l'on estime significatifs pour le décrire. »

Cependant, l'usage de synonymes demeure une stratégie d'explication employée non seulement par les apprentis-enseignants mais aussi par les apprenants. Chez les apprentis-enseignants, les synonymes se retrouvent seuls ou accompagnés d'autres stratégies d'explication.

7.3.2.3.1 Synonymes pour résoudre rapidement l'incompréhension d'un mot-clé

Un synonyme, comme une définition, peut être employé par l'apprenti-enseignant lors de l'explication de la consigne d'une activité. Dans l'exemple suivant, un synonyme est employé comme la seule stratégie d'explication pour résoudre l'incompréhension d'un mot inconnu :

Transcription de la parole		
1	SEV	alors\ est ce qu'il y'a des mots/ (0,3) que tu ne comprends pas\ (1,2) ou ça va\ (1,5)
2	NAO	mh: (0,6) pense bête/ (1,0)
3	SEV	un pense bête/ c'est un: mémo (0,5) tu sais/ (.) c'est [une li]ste/ (.) okay/ (0,5) [alors\]
4	NAO	[OH] (1,7) [oui]

Exemple 12 : Utilisation de deux synonymes

Séverine commence l'activité par demander à l'apprenante s'il y a des mots inconnus. Le besoin de connaître les mots-clés peut ainsi se faire ressentir (cf. chapitre suivant), et l'apprenante peut faire part d'une incompréhension. Un autre avantage d'aborder le vocabulaire en amont de l'activité est que la tâche devient moins couteuse cognitivement, et qu'il pourrait y avoir plus d'usage productif des mots-clés pendant l'activité (Skehan, 1996). Séverine utilise deux synonymes (ligne 3) pour résoudre rapidement l'incompréhension du mot-clé. Les deux synonymes qu'elle emploie sont transparents aux mots en anglais, ce qui augmente davantage l'efficacité. L'usage de synonymes est adapté à cette situation où la compréhension rapide des mots-clés est de première importance. L'usage de synonymes semble avoir été efficace car Naomi réagit vivement à la ligne 4.

La synonymie est utile pour résoudre une confusion potentielle entre deux mots qui sont phonologiquement proches[3] suite à un problème de prononciation. Ce type de confusion arrive très fréquemment : « Confondre des mots entre eux en raison de leur ressemblance formelle et/ou sémantique fait partie intégrante du processus d'apprentissage du vocabulaire d'une L2 » (Théophanous, 2001 : 117). Quand il y a le risque qu'une faute de prononciation soit liée à la méconnaissance d'une paire minimale ou d'une paire de mots qui se ressemblent phonologiquement, une explication simple et générale s'impose. Dans cet exemple, Mélissa a recours à la synonymie.

Transcription de la parole		
1	MEL	en espagne/ (0,2) ils travaillent (.) vous travaillez plutôt où/ alors/ (0,4) °est-ce° que vous [travaillez dans des] bars/ dans&
2	ALJ	[et]
3	MEL	&des
4	ALJ	normalement en les bars/ (.) et hum magazines/ (0,6)
5	MEL	[ouais/ (.) dans des m]aga[S]INS/
6	ALJ	[oh:\] (0,2) [mh] (0,6) magasins/ (0,5)
7	MEL	°puisque° magazine/ tu sais c'est les catalogues\ (0,5)
8	ALJ	AH:\ euh (0,5) [d'accord\]
9	MEL	[magazi]ne c'est les c'est les petits catalogues/ comme les journaux/ (.) et magaSIN/ (0,6) c'est euh: des boutiques (0,5)
10	ALJ	magasin\ (2,9)
11	MEL	et euh ouais/ (0,2) d'accord/

Exemple 13 : Utilisation de synonymes pour désambiguïser une paire minimale

[3] Les paires minimales ou des mots phonologiquement proches constituent un type de « mots difficiles » selon Tréville & Duquette (1996).

L'énoncé de l'apprenante à la ligne 4 ne nous permet pas de savoir s'il s'agit d'une simple faute de prononciation ou si elle est consciente de l'existence des deux mots qui se ressemblent. Afin de couvrir toutes les possibilités, Mélissa corrige la prononciation dans un premier temps en utilisant de la saillance vocalique (ligne 5) et donne ensuite des synonymes (lignes 7 et 9). L'apprenante répète le mot juste à la ligne 10 avec la bonne prononciation. Courts et simples, les synonymes semblent constituer une bonne stratégie d'explication quand l'apprenant doit garder en mémoire deux définitions.

La synonymie peut aussi être utilisée en combinaison avec d'autres stratégies d'explication (comme une mise en situation, une traduction, une définition, un geste) pour résoudre l'incompréhension d'un item lexical. Un synonyme est parfois plus facile à comprendre qu'une définition formelle, comme en témoignent les réactions de Liam quand Victor lui explique le sens des « congés payés ».

Transcription de la parole			Image du participant	Ph. observés
1	VIC	est-ce que tu sais ce que c'est/ (.) les congés payés\ (1,0) est-ce que tu connais l'expression (1,6)		Liam s'approche de l'écran comme il a l'habitude de faire quand il ne comprend pas
2	LIM	euh (1,9) non (0,4) non\ (0,5)		
3	VIC	les congés payés/ (.) **ça fait référence/ (.) aux jours/ euh de vacances/ (.) que les travailleurs/ que les employés/ (0,4) peuvent prendre/ (0,5) en étant payés (0,3) donc c'est euh les vacances (1,0)**		
4	LIM	**ah (.) oui**		Liam s'éloigne de l'écran et perd son expression d'incertitude

Exemple 14 : La synonymie pour compléter une définition

Si un synonyme peut être plus efficace qu'une définition formelle pour la compréhension, l'avantage d'utiliser cette combinaison est que pendant la définition, l'apprenant se fait des hypothèses, qui sont par la suite confirmées ou infirmées par le synonyme. Victor a l'habitude de donner des définitions formelles, lexicographiques et non adaptées au niveau langagier de son apprenant, suivies de traductions ou de simples synonymes. Les synonymes et traductions

employés par Victor portent à croire que les explications courtes et simples sont plus efficaces que les définitions longues et complexes (voir Chaudron, 1982 pour un constat similaire).

Parfois, le synonyme ne résout pas immédiatement la compréhension, ce qui pousse l'apprenti-enseignant à utiliser d'autres stratégies. Dans l'exemple suivant, Emilie emploie d'abord un synonyme, puis passe à une définition et à une traduction quand sa première stratégie ne fonctionne pas.

Transcription de la parole		
1	EMI	et euh tu vas faire le sapin cette année/ (0,3) vous avez déjà le sapin/ (1,6)
2	FIN	euh: (1,0)
3	EMI	le sapin [de] noël\ (.) l'arbre\
4	FIN	[se] (3,6) euh: (1,1) qu'est-ce que c'est (1,5) ((rire)) (0,3)
5	EMI	hum: c'est l'arbre de noël (0,4) c'est euh: [le christm]as tree&
6	FIN	[oh oh oui oui]
7	EMI	&((rire)) (0,6) [((rire))]
8	FIN	"oh yeah yeah" [((rire))] (.) hum (0,6) euh: oh oui

Exemple 15 : La synonymie comme première stratégie

Cet exemple illustre la capacité qu'ont les apprentis-enseignants à adapter leurs explications en fonction des réactions des apprenants. Si la synonymie ne fonctionne pas, ils passent à d'autres stratégies comme la définition ou la traduction. Ils peuvent aussi recourir à l'usage de gestes (cf. section suivante).

Dans les exemples précédents, les synonymes sont utilisés pour résoudre des incompréhensions manifestes. Un synonyme peut aussi être utilisé pour vérifier ou confirmer la compréhension de l'apprenant. Si les apprenants disent avoir compris le sens d'un item lexical sans l'avoir prouvé, un synonyme peut être employé pour être certain, comme le fait Adèle dans cet exemple.

Transcription de la parole			Image du participant	Ph. observés
1	ADL	euh je vois que vous l'avez compris/ parce que vous m'avez dit/ il y a/ un paradoxe\ (1,0) vous connaissez paradoxe/ (1,0)		Geste métaphorique, mouvement de main tenant le concept de contradiction
2	ALN	oui/ (0,3)		
3	CTR	°mh mh°/		
4	ADL	une **contradiction/ (0,3)** toi aussi catriona/ (0,7)		
5	CTR	oui		

Exemple 16 : Vérifier la compréhension par l'emploi d'un synonyme

Cet exemple est issu de la première semaine où les apprenants visionnent une vidéo qui décrit une entreprise qui, à l'instar d'une entreprise américaine, donne des congés payés illimités à leurs employés. La vidéo cite une étude américaine qui a trouvé que bien qu'ayant des congés illimités, les employés dans ce type d'entreprise prennent moins de vacances que les employés dans des entreprises classiques. Après que les apprenantes font part de leurs avis sur ce type d'entreprise, Adèle souhaite aborder ce paradoxe. Adèle est sûre que les apprenants ont compris la vidéo, car elles en parlent, mais elle ignore si les apprenantes connaissent le mot paradoxe. Pour relier la forme à son sens, Adèle emploie un synonyme accompagné d'un geste métaphorique (ligne 4). Les synonymes sont bien adaptés à vérifier rapidement la compréhension.

7.3.2.3.2 Confirmer l'hypothèse de l'apprenant et proposer une formulation plus adaptée

En plus de vérifier la compréhension, les synonymes peuvent être utilisés par les apprentis-enseignants pour confirmer les hypothèses des apprenants. L'exemple suivant est issu d'une activité où les apprenants doivent décrire une caricature qui s'affiche sur l'écran. Liam, incertain de l'existence du mot « contrôle » en français, le prononce à l'anglaise avec une intonation montante. Victor profite de ce moment d'apprentissage :

		Transcription de la parole
1	LIM	l'employeur est très grand/ (0,3) e[uh:]
2	VIC	[mh] mh
		(0,8)
3	LIM	parce que euh il avait euh beaucoup de: (1,6) hum (2,1) "control"/
		(0,7)
4	VIC	oui/ (.) oui oui (0,6) t[rès bie]n
5	LIM	[°et°] (1,2) hum (0,5) c'est le: (0,2) correct mot/ (.) non/
		(0,9)
6	VIC	oui/ oui oui c'est le le bon mot
7	LIM	oui °oui°/
		(0,5)
8	VIC	beaucoup de contrôle (0,3) beaucoup de pouvoir en fait

Exemple 17 : La synonymie pour confirmer l'hypothèse de son apprenant

Après avoir confirmé l'existence du mot en français, Victor ajoute un synonyme (ligne 8) afin que le mot que l'apprenant vient de tester s'installe et se solidifie dans un réseau sémantique. Victor pallie ainsi le manque de représentations en L2 dont parle Bogaards (1994). Un synonyme peut avoir la double fonction de servir de définition du mot juste et d'augmenter le nombre de mots nouveaux auxquels l'apprenant est exposé.

Cette fonction de confirmation est aussi valable pour les mots justes qui sont donnés par les apprentis-enseignants. Dans l'exemple suivant, Emilie donne un synonyme écrit afin d'éviter la surcharge cognitive quand elle donne un mot juste à l'oral et à l'écrit. Ainsi, l'apprenant peut noter les deux mots et les réviser plus tard.

		Transcription de la parole
1	EMI	est-ce que vous trouvez qu'il y'a des inconvénients (0,2) du coup (0,4) à travailler en équipe (2,1)
2	ADN	euh: (0,6) ce n'est pas (0,2) euh: (0,7) mal euh (0,7) il n'y a pas (.) beaucoup des inconvénients (0,2) euh (0,3) (°mh°) de temps en temps quelque fois (0,3) euh (.) il y a: (0,2) des: (1,7) "arguments" (.) arguments (.) euh
3	EMI	des dispu[tes]
4	ADN	[en-] (0,6) disputes et: entre: (0,4) euh: (0,9) les gens ont (.) marre (.) euh:
5	EMI	<((clavardage)) disputes, conflits> (0,8)
6	ADN	euh (.) c'est norma- (.) normal (0,5) mais euh: (0,6) euh (0,9) dans la plupart (.) de (0,5) il n'y pas des inconvénients (.) dans mon avis (1,5)
7	EMI	d'accord\ (.) merci

Exemple 18 : Donner un synonyme discrètement par clavardage

Cet exemple est issu d'une activité dont le thème est le travail en équipe. Emilie demande à ses apprenants s'il y a des inconvénients, ce qui met Aiden en détresse verbale (Lüdi, 1994) car il cherche le mot « disputes » mais ne trouve pas la forme en français. Cette panne lexicale se manifeste par l'usage de l'anglais suivi de l'emploi d'un faux ami pour le sens souhaité. Il s'agit de la recherche productive, ce qui est selon Laufer & Hulstijn (2001) le type de recherche le plus propice à l'acquisition du vocabulaire. Emilie donne le mot juste en français (ligne 3), et Aiden le répète immédiatement après (ligne 4). Emilie le répète par écrit (ligne 5), ce que les apprentis-enseignants ont l'habitude de faire quand ils donnent un mot inconnu (cf. chapitre précédent), mais elle va plus loin en donnant un deuxième synonyme à l'écrit. Les apprenants peuvent ainsi sélectionner leur niveau d'implication et sélectionner le mode par lequel ils apprennent. Cette définition par substitution (Flowerdew, 1992) est faite de manière peu intrusive afin de ne pas interrompre le fil de la communication, car l'objectif de cette question est d'engendrer l'output communicatif focalisé sur le sens. Emilie donne un deuxième synonyme uniquement par écrit (conflits). Le synonyme a l'avantage aussi de s'assurer que les participants parlent de la même chose au cas où le mot « disputes » n'est pas compris.

L'emploi de synonymes constitue donc une bonne manière de résoudre rapidement l'incompréhension, désambiguïser des paires minimales, compléter des définitions plus

complexes, vérifier la compréhension, confirmer des hypothèses des apprenants, et s'assurer de l'intercompréhension. Les apprentis-enseignants montrent un certain degré de conscientisation des stratégies d'explication qu'ils emploient, et la synonymie ne fait pas exception. Observons la façon dont Mélissa préconise l'usage de cette stratégie quand elle délègue la tâche d'explication à son apprenante la plus forte :

		Transcription de la parole
1	MEL	alors est-ce que v- est-ce que vous avez compris ce que sont les congés p- (0,7) les congés/ euh (0,3) tu sais ce que c'est aussi [alejandra/]
2	ALJ	[°oui°/] (1,0) je ne sais pas (0,3) les problèmes est-ce que j'ai étudié trans- euh (0,4) "translation and interpreting" je n'ai étudié "business" (inaud.) <((en riant)) je ne [veux p]as> [((rires))]
3	MEL	[ok\] (0,5) [pas de <((en riant)) pro]blème>\ y a pas de problème\
4	ALJ	((rires))
5	MEL	euh ben hum ana/ est-ce que tu peux expliquer à alejandra en français ce que sont les congés/ peut-être\ (0,4) un synon[yme/]
6	ANA	[euh] les con-/ (0,5) oui euh les congés sont comme euh les vacances/ (1,2)
7	MEL	voilà\ (0,3) parfait\ (.) [((rires))]
8	ALJ	[ah] ah okay\ (0,4)
9	MEL	d'accord/
10	ALJ	d'accord\

Exemple 19 : Mélissa préconise l'usage d'un synonyme

Après avoir introduit le sujet des congés payés en France en le contextualisant, Mélissa vérifie la compréhension des deux apprenantes individuellement car c'est l'un des mots-clés de la tâche. Alejandra s'excuse de ne pas connaître le mot (ligne 2) et Mélissa emploie Ana à l'aider (cf. chapitre suivant). Ce faisant, Mélissa promeut l'utilisation d'un synonyme (ligne 5), ce qui nous nous renseigne sur son niveau de conscientisation des stratégies d'explication et sur ses croyances quant aux meilleures. Ana donne un synonyme à la ligne 6, ce qui résout le problème d'incompréhension.

Pour résumer, dans le cas d'une incompréhension, les apprentis-enseignants emploient des synonymes comme la seule stratégie d'explication ou la combinent avec d'autres stratégies comme la mise en situation, la traduction ou la définition. Les synonymes sont employés aussi pour vérifier la compréhension de l'apprenant, confirmer l'hypothèse de l'apprenant, et pour se mettre d'accord sur le sens d'un mot juste. Peu de définitions et synonymes se font par clavardage, et aucun ne se fait uniquement par ce mode. Samia, par exemple, donne une définition des congés payés par écrit qui n'est qu'une répétition exacte de

sa définition orale[4]. La synonymie semble être une stratégie efficace non seulement pour résoudre rapidement l'incompréhension de mots inconnus, mais aussi pour étendre le lexique des apprenants, se mettre d'accord sur le sens d'un mot juste recherché, et pour créer des liens mentaux chez les apprenants. Le clavardage semble se prêter bien à donner des synonymes de manière peu intrusive sans interrompre le fil de la communication.

7.4 Explications ostensives

Dans la section précédente, nous avons analysé la façon dont les apprentis-enseignants emploient des stratégies voco-verbales lors des séquences d'explication lexicale et leurs diverses fonctions au cours de celles-ci. Ici, nous nous efforçons d'aborder la façon dont le sens est abordé de par des explications ostensives, c'est-à-dire des explications par gestes et images. Certains auteurs comme Bogaards (1994 : 170) rejettent ces procédés explicatifs comme étant des « moyens extra-linguistiques » : « gestes, mimes, images, films, objets réels, etc. ». Bogaards n'a pas confiance en ces méthodes d'explication, car il pense que ces ressources ne marchent que pour des éléments qui sont « perceptibles sous une forme visuelle ou auditive » (170) et qu'une représentation visuelle est susceptible d'induire l'apprenant en erreur. Il dit que :

> « Le matériel visuel semble justement pousser à une interprétation fondée sur la L1 (cf : Scherfer 1985 : 423), ce qui peut faire négliger les différences présentes dans les faits culturels. Il est pour cette raison sans doute préférable d'utiliser des photos soulignant ce contexte culturel, plutôt que des dessins, qui sont trop souvent culturellement « neutres » » (170).

Nous nous inscrivons en faux contre l'idée qu'un dessin soit culturellement neutre. A titre d'exemple, l'image suivante est très française en ce qu'elle touche à des enjeux qui sont d'actualité en France. De plus, elle provoque les séquences d'explication lexicale suivantes : « les 35h », « le Père Noël », et « la pause café », ainsi que d'autres items lexicaux connexes comme les congés payés.

[4] Cet exemple n'est pas transcrit ici.



Figure 4 : Exemple d'un dessin qui a été envoyé aux apprenants

Nous rejoignons McNeill (1992) qui pense que les gestes et la parole font partie d'un seul et même système cognitif, Mayer (2001) qui a démontré que les images et les mots ensemble est bénéfique pour l'apprentissage, et et Sindoni (2013) qui plaide pour que toutes les modalités soient prises en compte lors des interactions en visioconférence. Puisque la visioconférence offre en plus du microphone et du clavardage une webcam et le partage de documents, la façon dont ces deux derniers canaux sont employés par les apprentis-enseignants et combinés avec la parole orale lors des explications lexicales est digne d'intérêt. La communication en visioconférence est avantageuse car la webcam offre aux participants des ressources en plus qui sont absentes lors de la communication par audioconférence ou par clavardage uniquement. L'usage de gestes et de photos se résume comme suit.

	Adèle	Mélissa	Victor	Samia	Emilie	Étienne	Séverine	Total
Gestes	16	14	8	3	5	2		48
Photos	1	2	2	2				7
Total	17	16	10	5	5	2	0	55

Figure 5 : Nombre de séquences dans lesquelles les gestes sont utilisés pour se focaliser sur le sens

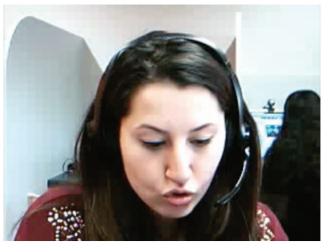
Dans plus d'une séquence sur six, des ressources visuelles sont utilisées pour se focaliser sur le sens des items lexicaux. Tous les apprentis-enseignants sauf Séverine utilisent leurs corps, et quatre sur sept utilisent des photos, du moins pendant les séances que nous avons transcrites. Les trois sous-sections suivantes vont aborder l'usage d'images, de sons, et de gestes pour expliquer le sens des items lexicaux.

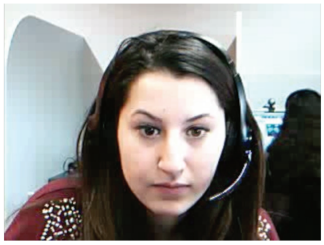
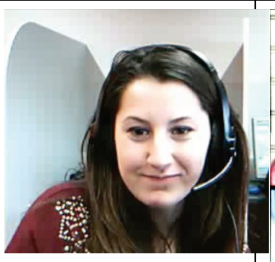
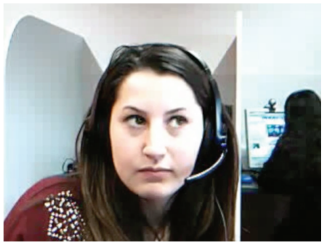
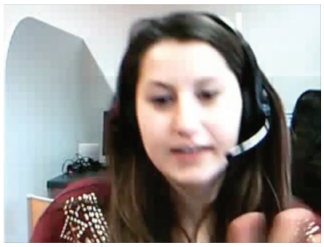
7.4.1 Usage d'images

Les images ont la particularité de ne pouvoir être spontanées comme la parole. En effet, Visu offre aux apprentis-enseignants la possibilité d'envoyer des images, des mots-clés et des vidéos aux apprenants durant la séance, mais ces supports doivent être téléchargés sur la plateforme en amont des interactions. L'envoi de chaque image a donc été planifié. La sélection d'images en amont de la séance peut avoir un impact à la fois sur les mots qui sont expliqués et sur la façon dont les explications se déroulent. Nous allons voir que l'image peut constituer une stratégie d'explication efficace, mais il y a toujours un risque que l'apprenti-enseignant soit trop dépendant du support audiovisuel au détriment d'une bonne explication.

7.4.1.1 Utilisation d'images pour remplacer une explication verbale

L'envoi d'une image peut constituer la seule stratégie explicative employée par un apprenti-enseignant lors d'une séquence d'explication. L'exemple suivant est issu de la cinquième séance où les participants doivent élaborer un menu pour un camion-restaurant. La sélection en amont de la séance de l'image d'une quenelle a comme effet premier que Samia s'attarde sur cet item lexical et non pas sur celui qui est adjacent (tarte aux pralines) :

Transcription de la parole	**Image du participant**	**Phénomènes observés**
1 ANG quelle sorte de repas (0,2) avez-vous: préparé dans: votre restaurant\ (1,6) 2 SAM [dans mon ancien res]taurant/ 3 ANG [c'était des^ha-] (0,5) ouais (0,5) 4 SAM euh c'était euh: des plats en fait euh lyonnais **(2,1) donc en** dessert y avait la tarte aux pralines/ **(7,0)** <((clavardage)) la tarte aux pralines> hu:m**: (0,4) en plat principal euh:** (0,5) en plat**: (1,5)** °**principal/**° (0,3) °en plat principal/° **(1,1) je (0,3) cuisinais** des quenelles/ **(1,3) vous savez c`que c'est** des quenelles/ (1,1) 5 ANG °**euh:**		Regards vers le clavier

6	SAM	<((clavardage)) en plat principal je cuisinais des quenelles> **(16,4) vous [sav]ez c`que&**		De longs regards vers écran en attendant une réponse des apprenants
7	SEN	**[le]**		
8	SAM	**&[c'est des quenelles]/**		
9	SEN	**[°(inaud.)°]**		
10	ANG	**[euh:] (0,2) oui (0,9) c'est avec (0,3) de la soupe/ (4,5)**		
11	SAM	**non/ (10,7)**		
		<((image))>		Samia envoie une image d'une quenelle
		(1,3) j`vous ai envoyé une image/ pour que vous puissiez voir c`que c'est une quenelle\ **(6,4) ((rire))** (11,4)	 	Samia sourit légèrement après avoir envoyé l'image, et les apprenants regardent l'image sur leurs écrans
12	SEN	d'accord (0,3) les comme [les]&		
13	SAM	[d'accord/]		
14	SEN	&euh: (1,3) pommes de terre/ (0,6)		
15	SAM	oui (0,3) euh: ouais **(0,4)** sauf que c'est comme euh: **(0,3) c'est pas vraiment des pommes de terre** hein c'est pas des pommes de terre du tout même/ (0,4) c'est hum: **(1,7)** des gros pff c'est pas des pâtes/ **(1,4)** hu:m (0,3)		Regards pensifs
		`fin bon ce ça **ressemble** à ce qu'il y a sur l'image\ (0,8) ((rire))		Geste déictique peu visible vers l'écran

Exemple 20 : Une flexibilité multimodale limitée

A la ligne 4, Samia dit avoir servi des tartes aux pralines et des quenelles, et répète ces items lexicaux par clavardage. Ensuite, elle regarde l'écran pendant 16,4 secondes en attendant que les apprenants réagissent. Elle vérifie spécifiquement la compréhension de la quenelle mais décide de ne pas s'attarder sur la tarte aux pralines. On peut se demander si la présence de l'image a orienté ce choix didactique. Samia semble ne pas avoir entendu la réponse d'Angela

(ligne 10) et elle continue à regarder vers l'écran pendant plus de 10 secondes. Après la longue pause, elle envoie une image (ligne 11) et commente cette action. 17 secondes passent avant que Sean ne réagisse, et il vérifie sa compréhension en disant que c'est comme des pommes de terre. Samia semble avoir voulu se contenter de l'image et ne pas avoir été prête à expliquer le sens d'une quenelle verbalement. Suite à son échec d'explication (ligne 15), Samia verbalise, à l'aide d'un geste peu visible, que l'image offre la meilleure description du sens (« `fin bon ce ça ressemble à ce qu'il y a sur l'image »). On peut dire alors que l'explication est monomodale car le sens n'est véhiculé que par l'image. Si l'image n'est pas reconnue, ce qui peut être le cas pour qui ne connaît pas la cuisine française, ce support est complètement inutile. On voit bien que la préparation d'une image ne dispense pas la préparation d'une explication langagière.

Ce qui est intéressant dans cet exemple d'un point de vue multimodal est que le canal des images prend une haute densité modale (Norris, 2004) quand l'image est envoyée. Rien n'est dit pendant près de 18 secondes suite au commentaire méta-sémiotique de Samia (ligne 11). Pendant ce temps, l'image occupe une place primaire, et est commenté, et atteint ainsi une haute intensité modale (cf. chapitre 3). Pour rappel, un mode devient hautement intense si :

- D'autres modes sont structurés par celui-ci ;
- L'action de haut niveau réalisée par l'acteur à l'aide de ce mode hautement intense serait affectée et/ou rendue impossible si ce mode n'était pas hautement intense ;
- Tous les « embodied modes » sont focalisés sur ce mode (Norris, 2004).

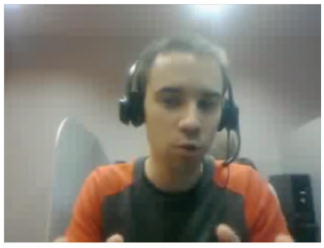
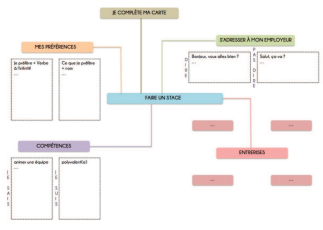
Le statut du mode utilisé par Samia pour envoyer la photo remplit ces trois conditions. Premièrement, le mode voco-verbal est affecté comme en témoignent le commentaire méta-sémiotique de Samia (« je vous ai envoyé une image », ligne 11), le long silence qui suit et la discussion sur l'image. Deuxièmement, l'explication de la quenelle serait impossible sans l'image, puisque Samia est incapable de la décrire verbalement (ligne 15). Troisièmement, le geste de Samia (ligne 15) et les regards des apprenants (ligne 11) sont tous orientés vers ce mode.

D'un point de vue cognitif, le processeur visuel du cerveau est alimenté par l'image, tandis que le processeur auditif reste inutilisé car Samia ne donne pas de définition verbale (Mayer, 2001). Cette explication n'est donc pas très bien conçue pour exploiter au maximum les ressources cognitives de ses apprenants.

D'un point de vue pédagogique, il semble que la préparation de cette image ait orienté les choix didactiques de l'apprentie-enseignante quant à la focalisation sur les items lexicaux. Elle a choisi de s'attarder sur la quenelle mais non pas sur la tarte aux pralines. On se demande si elle se serait attardée sur cette dernière si elle en avait préparé une photo, car l'utilisation d'images n'est possible que pour les items lexicaux dont l'explication a été anticipée en amont. Pour expliquer par photo un item lexical dont l'incompréhension n'a pas été anticipée, il faudrait effectuer une recherche sur Internet et envoyer le lien. Ce serait similaire à l'acte de dessiner au tableau dans une salle de classe. On pourrait imaginer une plateforme qui permette aux participants de dessiner sur l'écran, mais ce n'est pas une affordance de Visu. Il semble ici que la préparation de l'image se soit faite au détriment d'une préparation explicative verbale. Tout porte à croire que l'image a permis à l'enseignante de s'exempter d'une explication verbale.

7.4.1.2 Utilisation d'images en combinaison avec d'autres stratégies explicatives

L'envoi d'images peut aussi être une stratégie d'explication qui se combine avec d'autres. Dans l'exemple suivant, Victor explique le sens d'une « carte mentale » avant d'en envoyer une représentation visuelle :

Transcription de la parole	**Image du participant**	**Phénomènes observés**
1 VIC est-ce que tu sais c`que c'est une carte mentale/ (1,5) 2 LIM euh: non je ne comprends: (0,3) pas: le mot (0,4) carte mentale/ (0,5) 3 VIC ah\ (0,4) euh: je sais pas du tout c`que c'est en anglais\ (.) une carte mentale/ c'est une façon/ c'est un schéma/ (0,9) un **tableau/ (.) qui te permet** de: hum: (.) résumer/ (0,3) des informations\ (0,3) je vais t'envoyer la carte mentale/ (0,2) comme ça peut-être que tu pourras euh voir/ de quoi ça a l'air\		Victor fait un geste iconique qui est quasiment 100% hors champ
<((image))		Victor envoie une image d'une carte mentale

dans le chat (22,7)		

Exemple 21 : Une flexibilité multimodale élevée

De même que Samia dans l'exemple précédent, Victor envoie une image et émet un commentaire méta-sémiotique (« dans le chat », ligne 3). Suite à l'envoi de l'image, ce canal devient hautement dense et intense, car il n'y a une longue pause de 22,7 secondes pendant laquelle l'apprenant se concentre sur l'image. Les similitudes s'arrêtent là. Contrairement à Samia, qui n'explique pas le sens de l'item lexical représenté par l'image, Victor l'explique d'abord avant d'envoyer la représentation visuelle. Victor fait montre d'une plus grande flexibilité sémiotique, multimodale et pédagogique, car il est capable d'expliquer le sens de l'item lexical non seulement par l'envoi d'une image mais aussi par une explication langagière. Cependant, Victor laisse passer une opportunité d'utiliser à la fois le processeur visuel et le processeur auditif (Mayer, 2001) de Liam en envoyant la photo *avant* son explication verbale.

En ce qui concerne la conscientisation de Victor de son utilisation des ressources sémiotiques, un déséquilibre s'observe. Quand Victor explique le sens de la carte mentale (ligne 3), il fait un geste iconique pour représenter un tableau. La plus grande partie de ses mains sont hors champ, ce qui rend quasiment invisibles ses gestes. Il semble être plus conscient de son utilisation de l'image, comme en témoigne son commentaire méta-sémiotique, qu'il ne l'est de son utilisation de gestes. On peut faire l'hypothèse que l'utilisation de ce dernier mode sémiotique, dont il se sert *a priori* tous les jours, est devenue en quelque sorte banalisée, de façon à ce qu'il en ait perdu conscience. Par contre, le mode sémiotique par lequel il peut envoyer des images constitue un mode dont l'utilisation, du moins en début de sa carrière d'enseignant en visioconférence, est réfléchie, planifiée, conscientisée. L'envoi d'une photo demande un plus grand effort que la production d'un geste qui se fait naturellement. Il revient donc aux formateurs et aux apprentis-enseignants de reprendre conscience de tous les modes sémiotiques dont ils se servent lors de l'activité pédagogique, y compris les modes banalisés comme la voix et les gestes (Gadoni & Tellier, 2014), et de les remettre au même niveau de conscientisation que les modes novateurs ou rarement utilisés comme l'envoi d'images.

Dans cette section, nous avons montré que l'envoi d'images a le potentiel de constituer une stratégie pédagogique bien adaptée à l'explication lexicale à condition que l'apprenti-enseignant s'arme d'une certaine flexibilité multimodale et pédagogique. A l'instar de Cicurel (2005), qui élabore la notion de *flexibilité communicative*, nous pouvons dire que Victor fait

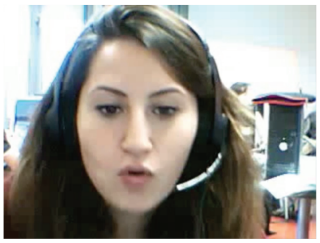
montre d'une plus grande *flexibilité multimodale* que Samia dans les deux exemples ci-dessus. Tandis que Samia n'est capable d'expliquer le sens du mot « quenelle » que par la photo, Victor explique le sens du mot « bilan » visuellement et oralement. L'image ne devrait en aucun cas remplacer une explication langagière, mais devrait être plutôt envisagée comme un support complémentaire. L'envoi d'images, en raison de la planification nécessaire et à l'effort requis pour l'envoyer, semble être l'une des stratégies d'explication les plus conscientisées. A notre connaissance, chaque envoi d'une photo s'accompagne d'un commentaire méta-sémiotique.[5] La préparation en amont de la photo devrait donc s'accompagner d'une réflexion aussi conscientisée sur la meilleure façon de délivrer une explication orale accompagnatrice.

7.4.2 Usage de sons

Le port d'un micro-casque fait en sorte que le canal voco-verbal n'a pas les mêmes caractéristiques que la communication voco-verbale en présentiel. Si le matériel fonctionne correctement, chaque bruit de bouche est perceptible, saillant, voire amplifié. Les apprentis-enseignants peuvent tirer profit de cette spécificité à des fins pédagogiques.

7.4.2.1 Utiliser les affordances de la visioconférence pour transmettre le sens d'un item lexical

Samia profite de la spécificité susmentionnée lors de son explication du mot « détente ». Cette apprentie-enseignante peine à provoquer une réponse de ses apprenants quand elle leur demande de penser à des mots, adjectifs et expressions à associer à « la pause café ». Pour leur donner une idée, elle dit que c'est un moment de détente, ce qui démarre une séquence d'explication autour du mot « détente ».

Transcription de la parole	**Image du participant**	**Phénomènes observés**
1 SAM pour moi c'est un moment d` détente par exemple (3,3) 2 ANG un moment de: quoi/ 3 SAM de DÉTENTE (0,5) où on se détend où on est euh (0,6) **((soufflement))** où on souffle (1,5) 4 ANG [d'accord] 5 SAM [où on a]rrête de travailler\ (1,6) où on se repose (2,5) un moment de détente		Samia souffle dans son micro qui amplifie le son

Exemple 22 : Explication ostensive sonore

[5] Depuis la disparation définitive de Visu, les commentaires méta-sémiotiques sont les seuls indices qu'une photo a été envoyée.

Quand Samia explique l'acte de se détendre, elle souffle dans son microphone, ce qui capte et amplifie le bruit de son souffle. Ce bruit est plus fort et plus perceptible qu'il ne serait dans une salle de classe, à moins d'avoir un microphone. La visioconférence se prête ainsi convenablement à ce type de démonstration sonore.

7.4.3 Usage de gestes

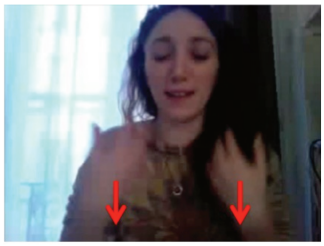
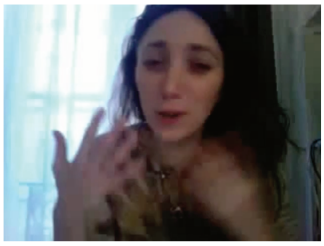
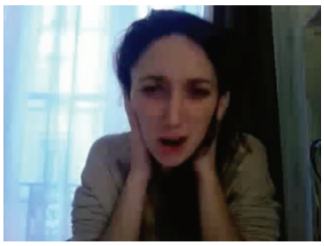
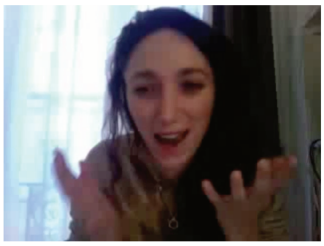
Malgré le fait que beaucoup de gestes soient produits hors champ, les gestes qui sont visibles sont souvent utilisés pour expliquer. Nous avons mis en évidence le fait que les gestes sont souvent rendus visibles lors des explications lexicales dans le chapitre 5 en page 160. On retrouve les trois fonctions de gestes pédagogiques décrits par Tellier (2008a) (informer, évaluer, animer) (cf. chapitre 5, section 5.4.2.4.3 en page 162) lors de toutes les phases des séquences d'explication (ouverture, noyau, clôture). Les gestes sont produits aussi bien par les apprenants que par les enseignants. Ici, nous nous attachons à décrire les différentes fonctions de ces gestes lors des explications lexicales en insistant sur leur importance.

7.4.3.1 Des gestes pour illustrer un concept concret

Les gestes illustrent des concepts concrets, comme un tableau (cf. exemple 21 en page 243) et des actions (cf. exemples 4 et 25). Quand Victor explique le mot « grève », son corps devient plus animé et il décrit gestuellement les verbes principaux de sa définition. Son geste est redondant pour Victor, qui comprend le sens du verbe « scander », mais pour Liam, il y a de fortes chances que le verbe soit incompris, et que le sens soit véhiculé uniquement par le geste. La webcam semble permettre aux apprentis-enseignants de livrer des explications renforcées de gestes, ce qui corrobore les résultats de Tellier & Stam (2012) qui ont trouvé que de futurs enseignants de FLE font des gestes plus amples, plus iconiques et qui durent plus longtemps quand ils expliquent des verbes d'action à des non natifs. Néanmoins, nous avons constaté lors de notre mémoire de Master 2 (Holt, 2014) qu'en visioconférence, les apprentis-enseignants ne semblent pas produire plus de gestes lors des séquences d'incompréhension.

7.4.3.2 Des gestes pour illustrer un concept abstrait

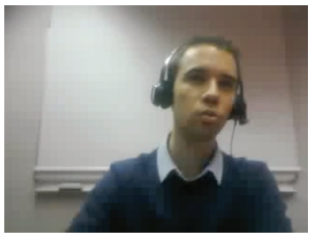
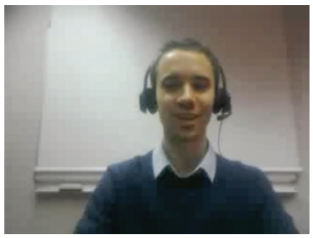
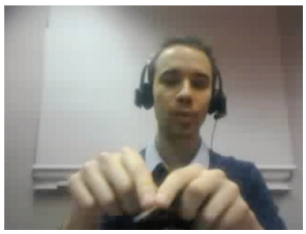
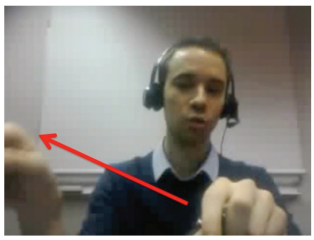
La description d'un concept abstrait risque d'être difficile en langue étrangère, mais les gestes peuvent rendre l'explication plus compréhensible. Les gestes faciaux peuvent expliquer des émotions, comme le fait Adèle dans l'exemple suivant.

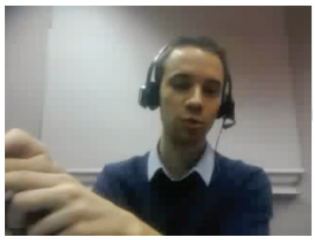
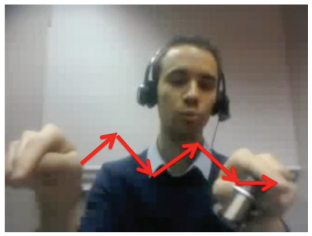
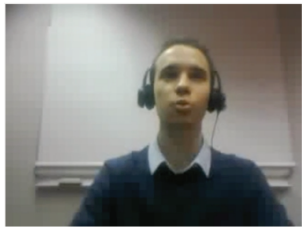
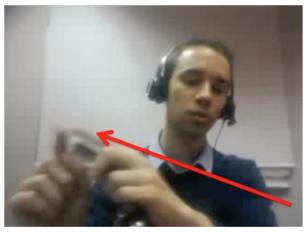
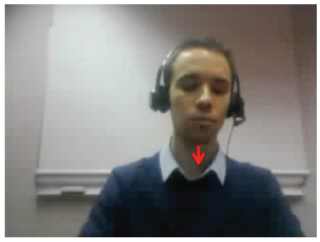
Transcription de la parole			Image du participant	Phénomènes observés
1	ADL	tu as l'air très très: décontenancée (0,3) décontenancée\		Paumes vers la poitrine, mouvement vers le bas
2	ALN	((rire)) (0,2)		
3	ADL	dé (.) con (.) te (.) nan- (0,5) -cée (0,4) <((clavardage)) décontenancée> (0,3) **c'est:** (1,1) hum (0,7) tu connais/ (0,7)		
4	CTR	oui je pense que oui oui ((rire))		Paumes vers la poitrine, mouvements dans tous les sens
5	ALN	dé (.) cont (.) ton (.) cée- [((rire))]		
6	ADL	[on sent que tu es:] **(0,3) euh**		
		(0,3) perdue déçue **euh o- oh: j'entends**		Adèle serre la tête avec ses mains
		pas (0,7)		Mains ouvertes, paumes vers le haut
7	CTR	un peu perdue oui\		

Exemple 23 : Usage de gestes et d'expressions faciales pour expliquer une émotion

Adèle utilise le mot « décontenancée » (ligne 1) et anticipe tout de suite une incompréhension. Elle le répète oralement et par écrit (lignes 1, 3) de manière autonymique pour le décontextualiser et faire de lui l'objet d'une explication. Pour expliquer le sens, elle donne deux synonymes, ainsi qu'une phrase exemple de quelqu'un qui est décontenancé (ligne 6). Cette définition orale est accompagnée d'expressions faciales et de gestes expressifs pour compléter la parole orale. Le sens est donc véhiculée par deux modalités à la fois : la parole et les gestes. Puisque le visage est tout près de la webcam, les expressions faciales sont perceptibles en permanence. Les apprentis-enseignants ont donc intérêt à tirer profit de cette ressource qui joue un rôle important en visioconférence (Cosnier & Develotte, 2011) et d'en développer une certaine conscientisation.

L'usage de gestes est utile pour expliquer des concepts abstraits qui seraient d'ailleurs difficiles d'appréhender par des explications purement orales. Dans l'exemple qui suit, Victor utilise ses gestes pour expliquer le sens d'une « rétroplanification ».

Parole orale	**Image participant**	**Phénomènes observés**
1 VIC tu crois euh par exemple est-ce-que c'est **mieux hum** est-ce-que tu sais ce que c'est/ une rétroplanification (2,6) 2 LIM hum: (2,9) 3 VIC <((clavardage)) Une rétroplanification.)>		Regard pensif lors d'une vérification de compréhension, puis regard clavier
(3,4) 4 LIM **euh: qu'est-ce que c'est une (0,3) rétroplanification/** (0,5) 5 VIC euh **tu sais pas ce que** c'est/ (.) tu n'as pas **une petite idée/** (1,6) 6 LIM **hu:m (4,7) un: (0,8) euh: (0,6) un chose** euh: (0,6) vieux/ (0,7) vieux/ (1,0)		Liam regarde vers la fenêtre de clavardage
7 VIC euh: non ah **oui c'est vrai y a le mot euh rétro (0,6) [euh: non en]** 8 LIM [oui ((rire))] 9 VIC **fait c'est quand:/ (0,3)**		Victor sourit
10 VIC euh: tu planifies un évènement/ (0,6) mais hum: en partant/ **au lieu de partir par exemple d'aujourd'hui (0,8)**		Victor donne une définition en français à l'aide de gestes
ah: jusqu'à la date euh: de l'évènement/ (.)		

	tu fais l'inverse tu pars de: la date de l'évènement/ (0,7) et tu fais		
	ta planification/ (0,2) à rebours (1,8)		
	donc **par exemple si j'organise quelque** re- que: **quelque chose pour le vingt-cinq euh:** (1,1) euh: décembre/ pour noël/ (0,3) j`décide euh: j`me dis bon ça m`prend euh: (0,3) les deux derniers jours je dois préparer: ça ça ça/ (0,6) **hum: (0,4) dix jours avant: je** dois avoir acheté tous mes: cadeaux/ (0,4) et cetera et cetera\		Regard pensif
	tu fais une planification mais en **partant euh:** (0,4) en faisant le ch`min inverse finalement\ (3,0)		L'index de la main gauche de Victor rejoint l'index de la main droite, déjà tendu
11 LIM 12 VIC 13 LIM 14 VIC	euh: (0,5) je pense que: euh: (0,3) je comprends (9,5) **d'ac[cord]** [euh:] le mot (0,2) oui (0,2) je pense (0,6) et dans l`cas d'un goûter/ (0,3) est-ce que tu sais euh: quelles seraient les étapes importantes/ si y fallait p- rétroplanifier (.) un goûter euh d'anniversaire/		Liam donne un indicateur de compréhension faible. Victor hoche la tête pour évaluer sa compréhension

Exemple 24 : Usage de gestes pour expliquer un concept abstrait

Au début de la séquence, Victor vérifie la compréhension de l'item lexical « rétroplanification » et l'inscrit dans la fenêtre de clavardage. Victor encourage le rappel réceptif (Laufer & Hulstijn, 2001 ; Nation & Webb, 2011) (ligne 5), et la répétition par écrit permet à l'apprenant d'étudier la forme, repérer les composants et faire des hypothèses. On

voit sur l'image (lignes 4-6) que Liam étudie la forme écrite du mot en orientant son regard vers la fenêtre de clavardage. Victor sourit et valorise l'attention que porte Liam aux composants du mot et son hypothèse (ligne 7). L'usage du clavardage au début d'une séquence est avantageux pour encourager le rappel réceptif et pour aider le repérage de la forme, ce qui est nécessaire avant que l'item lexical ne puisse être appris (Groot, 2000).

Quand Victor explique le sens du mot « rétroplanification », il met ses mains clairement dans le champ de la webcam (ligne 10), la ligne horizontale entre ses mains représentant l'axe temporel. Son indexe de gauche représente le présent et celui de droite représente la date de l'évènement, et avec ses indexes, il explique la relation entre le présent, la date future, et le processus par lequel on planifie un évènement à rebours en partant de la date de l'évènement. Suite à sa définition générale, Victor donne une définition liée à la tâche où il prend comme exemple le goûter d'anniversaire qui doit être planifié. Pour vérifier la compréhension, Victor sollicite une réponse appropriée à la tâche (Smith, 2003) en lui demandant les étapes nécessaires ainsi que les délais.

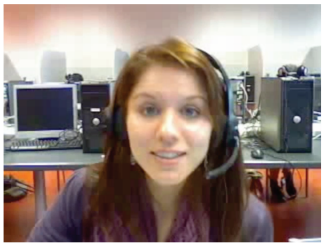
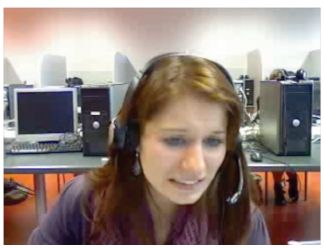
Dans cet exemple, le geste et la parole co-construisent l'explication et sont parfaitement synchronisés (Kendon, 2004). Les gestes se succèdent de façon à ce que la compréhension du deuxième geste dépende de la compréhension du premier. Quand Victor dit « au lieu de partir d'aujourd'hui jusqu'à la date de l'évènement » (ligne 10), il pose sa ligne horizontale. La direction de son geste est inversée lors de la partie suivante, quand il dit « tu fais l'inverse tu pars de la date de l'évènement et tu fais ta planification à rebours ». La représentation visuelle que donne Victor avec ses mains illustre clairement le fait que l'ordre soit inversé pour une rétroplanification. Le geste qui accompagne la phrase « à rebours » montre que selon Victor, il y a à peu près cinq étapes, à titre d'exemple. Le geste véhicule ainsi des informations qui ne sont pas contenues dans la phrase orale.

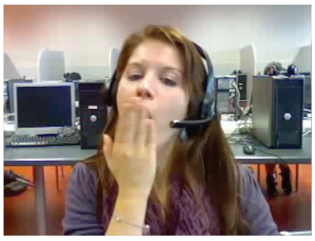
Les gestes de Victor sont très bien cadrés, car on les voit dans le centre de l'écran. On peut donc dire qu'ils sont conscientisés. Pour Victor, la planification à rebours représente un mouvement de droite à gauche, ce qui serait l'inverse d'une planification sur une ligne de temps classique de gauche à droite. Or Liam voit l'inverse. On ignore si Victor est conscient de ce phénomène, qui nécessiterait un niveau de conscientisation supérieur. Une autre particularité de la webcam est qu'ici, nous n'avons que trois images par seconde. Par conséquent, chaque flèche rouge correspond à une image, et on ignore si les mouvements de Victor ont été plus fluides. Une conscientisation approfondie des particularités de la communication par visioconférence, y compris une conscientisation des effets induits par le

faible nombre d'images par secondes, permettrait aux apprentis-enseignants de veiller à ne pas produire des mouvements trop petits ou rapides.

7.4.3.3 Gestes pour vérifier sa propre compréhension et pour décrire le mot juste recherché

La visioconférence se prête bien à la communication visuelle, à condition que les gestes soient produits dans le champ et que les participants regardent l'image de l'autre. L'exemple suivant illustre bien ce phénomène. Fasel Lauzon (2014 : 30) explique que l'explication est « non seulement un lieu de construction de connaissances, mais également un lieu de vérification de sa propre compréhension (pour la personne qui explique) […] ». Quand un apprenant demande la forme d'un item lexical, l'apprenti-enseignant doit être capable dans un premier temps d'avoir une compréhension adéquate de ce qui est recherché. Si l'apprenant n'est pas à même de faire du bricolage lexical (Lüdi, 1994) et décrire le mot juste qu'il recherche en français, et si une traduction est inutile faute de compétences en langues communes, un geste est susceptible d'être la meilleure façon de signaler une panne lexicale décrire ce que l'on cherche à dire (Kupetz, 2011). L'usage de gestes est un outil de grande valeur car il permet aux participants de communiquer par la webcam sans recours à aucun code linguistique. L'exemple suivant nous permet d'observer l'usage de gestes à ces fins par l'enseignante aussi bien que par l'apprenante. Dans cet extrait, Mélissa est en train de clôturer la séance quand Ana lui demande la traduction d'un mot inconnu.

Transcription de la parole	**Image de Mélissa**	**Phénomènes observés**
1 MEL euh bon [ben voilà/] ben& 2 ANA [oh et:] 3 MEL &écoutez (.) oui/ > **(0,4)** 4 ANA je peux te poser une question/ 5 MEL oui/ (0,8) 6 ANA oui euh comment est- [ce qu'on] d[it "YAW]NING"\ 7 >MEL [oui/] **(0,3)** [°((rires))°\]		Regards caméra
(2,6) **comment dit quoi/** (0,8) 8 ANA "yawning"/ (1,9) 9 MEL tu tu peux m'écrire le mot/ peut-être\ (4,3) 10 ANA <((clavardage)) yawning> (0,6) "I don't even know if it's spelled like that" (1,7) hum (2,4)		Visage d'incertitude

11 MEL **c'est un mot (0,2) anglais/ (0,9)** 12 ANA ouais\ (0,3)		Ana mime l'acte de bâiller
13 MEL AH 14 ANA x euh 15 MEL bâiller (0,7) 16 ANA bâiller\ 17 MEL qu- c'est quand tu fais **(1,1)** [comme ça/] 18 ANA [ouais ouais] (0,6)		Mélissa mime l'acte de bâiller
19 MEL c'est quand tu fais **(1,1)** okay (.) [((rires))]& 20 ANA [ouais ouais ouais\] 22 AJL [AH:\] 23 MEL <((en riant)) (alors)> [((rires))] [c'est le mot/]& 24 ALJ [((rires))] [merci] ((rires)) 25 MEL &ah ben peut- tu le connais/ alejandra/ (1,3) 26 ALJ [non je ne connais]& 27 MEL [tu connais le mot alejandra/] 28 ALJ &mais mais ["you"] expliques& 29 MEL [non] 30 ALJ &((rires)) (0,7) "you": (0,5) e[uh:] 31 MEL [ah (.)] allô/ 32 ALJ tu m'expliques euh (0,3) ((rires)) (0,5) 33 MEL °d'accord° (0,4) c'est <((clavardage)) bailler> c'est le mot/ (.) bâiller\	 	Mélissa répète son premier geste et fait des étirements tout de suite après

Exemple 25 : Utilisation de gestes pour trouver un mot juste

A la ligne 6, Ana demande la traduction d'un mot anglais que Mélissa ne comprend pas. Mélissa lui demande de l'écrire (ligne 9), mais cette stratégie de clarification ne permet toujours pas à Mélissa d'atteindre la compréhension nécessaire pour donner le mot juste. A la ligne 11, Ana fait un geste pour illustrer le sens du mot qu'elle cherche, et Mélissa, aussi par un geste, vérifie sa propre compréhension (ligne 17). Après avoir vérifié sa propre compréhension, Mélissa donne le mot juste à l'oral et à l'écrit (ligne 33) et le présente par le logonyme « mot ».

Ce qui est digne d'intérêt ici est la façon dont le geste est employé par Mélissa, car son geste à la ligne 19 obtient le statut autonymique. Pour certains auteurs comme pour Rey-

Debove (1997 : 83), les systèmes sémiotiques non sonores et non graphiques ne font pas partie de la langue mais sont des autonymes sémantiques :

> « On peut, à l'extrême limite, envisager des systèmes sémiotiques non sonores ou non graphiques, dans la situation de paralangages (gestuelle, etc.). Tous ces items peuvent être dénommés (ou présentés hors contexte) dans le discours de L_1 multisémiotique et inacceptable pour L_1. »

Pour cette auteure, l'introduction de ces ressources sémiotiques constitue une « interruption » qui n'est pas naturelle : « On peut imaginer un discours oral interrompu par un geste qui cite un item d'une gestuelle, et qu'on essaiera de transcrire artificiellement ainsi :

/Il a fait (haussement d'épaules) sans relever l'allusion/ » (84).

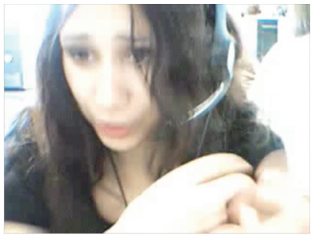
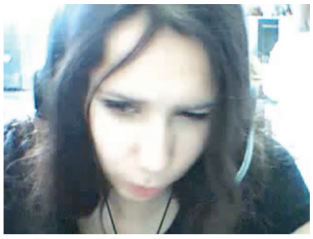
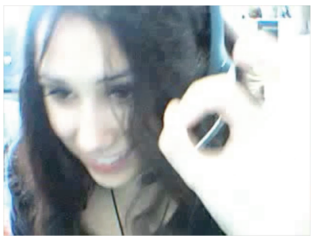
La parole de Mélissa à la ligne 17 serait alors transcrite ainsi : « /C'est quand tu fais (geste pour bâiller) comme ça/ », le geste étant autonymique. Alors que l'usage de gestes constitue une interruption pour Rey-Debove, nous rejoignons Sindoni (2013) lorsqu'elle argumente que toutes les ressources sémiotiques ont *a priori* le même statut et la même importance. Nous avons donc affaire au *mode-switching* (Sindoni, 2013) (alternance des modes ou alternance modale), plutôt qu'à une interruption.

Aussi digne d'intérêt est la densité modale qu'obtiennent les gestes de Mélissa aux lignes 17 et 19. Son début de phrase « c'est quand » suit au pied de la lettre le template pour une mise en situation (Fasel Lauzon, 2014), mais ce qui suit est un geste au lieu d'une explication linguistique. La communication sonore s'arrête complètement lors du geste, ce qui fait que sans ce geste, la communication n'existerait point. Voilà la définition du maximum degré de densité modale (Norris, 2004). Mélissa semble être bien consciente de son usage de gestes car son regard vers l'écran indique qu'elle vérifie le cadrage de ses gestes. Les gestes de Mélissa sont indissociables à sa parole, ce qui ne constitue pas une interruption, loin s'en faut. On assiste à un usage tout à fait naturel du corps de l'enseignante pour compléter sa parole, et la webcam facilite la transmission du sens. L'usage de gestes permet à Ana de se faire aider et à Mélissa de vérifier qu'elle donne l'item lexical recherché. La webcam permet aux apprenants de changer de mode facilement et manière fluide, quitte à produire des gestes autonymiques.

7.4.3.4 Utilisation de gestes pour relier le sens littéral à un sens figuré

Souvent, l'item lexical incompris comporte un sens littéral et un sens figuré. Afin d'approfondir la compréhension de l'item lexical, les apprentis-enseignants s'attachent à expliquer les deux sens par divers moyens. Après une tentative d'impliquer les apprenants (cf. chapitre suivant), Emilie explique le sens littéral ainsi que le sens figuré de l'expression

« arrondir les fins de mois » et les relie. Emilie explique non seulement le sens du verbe « arrondir » mais aussi le sens de toute l'expression.[6]

Transcription de la parole			Image du participant	Phénomènes observés
1	EMI	alors\ arrondir:/ euh: c'est quand tu prends quelque chose: un **objet/ (0,5)** et/ tu lui donne une forme ronde\ (1,5) MAIS/ (0,2) euh:\ là/ c'est pas: hm: c'est au sens figuré si tu veux donc tu le trouveras pas si je te donne le sens euh\ (0,7) le sens littéral (0,5) le le sens euh de base\ (.) du verbe arrondir\ (0,7) en fait/ (0,4) arrondir les fins de mois/ (0,6) ça veut dire gagner un peu plus d'argent/ que ce que tu gagnes déjà\ (1,2)		Geste iconique représentant un objet rond
		donc\ si par exemple tu touches une une bourse/ tu comprends/ **vous comprenez bourse/** (0,6)		Visage d'incertitude
2	FIN	oui (0,7)		
3	EMI	voilà\ (0,7) vous touchez une bourse/par exemple vos parents vous aident/ (0,4) mais vous voulez (.) [ARROND]IR les fins de mois (0,2)&		
4	FIN	[ah: oui\]		
5	EMI	&donc/ vous allez travailler\ (0,4) ça veut dire/ que vous voulez gagner/ (.) PLUS d'argent\ (1,9) parce que en fait/ en france/ (.) euh: je je ne sais pas comme c'est/ en irlande/ (0,5) mais en france/ (.) on reçoit le salaire et le paiement/ (.) à la fin du mois\ (0,5) donc/ (.) quand on ARRONDIT la fin du mois/ (0,6) on est sensé rajouter **UN ZÉRO/ (.) d'accord don[c avoir]**&		Geste emblématique représentant un zéro
6	ADN	[hm\]		
7	EMI	&une paye\ un peu plus grosse (.) un peu plus im[portante\]		
8	ADN	[oui oui oui]		
		(0,5)		
9	EMI	c'est pour ça/ qu'on dit (.) arrondir les fins de mois en fait/ (0,5) on arrondit les chiffres (0,6) sur le sur [le]&		
9	ADN	[hm\]		

[6] L'ouverture de cette séquence d'explication lexicale se trouve dans l'exemple 1 du chapitre 8 en page 262.

10 EMI	&chèque ou/ euh sur la carte bancaire\ (.) vous voyez/ (1,5) c'est une expression (0,3) voilà\ assez courante\		

Exemple 26 : Relier le sens littéral au sens figuré

Emilie commence son explication par donner une définition lexicographique et décontextualisé du verbe « arrondir ». Afin de renforcer le transfert du sens, elle maintient un geste iconique qui représente un objet rond devant la webcam et appuie sa définition d'une pause longue de 1500 millisecondes. La geste nous renseigne sur la forme de l'objet avant la parole. Après, Emilie dit que le sens littéral ne va pas permettre aux apprenants de deviner le sens figuré.[7] Pour expliquer le sens figuré, Emilie utilise un nouveau geste, cette fois-ci un geste qui représente un zéro qui serait ajouté au nombre indiqué sur une fiche de paie. Le geste iconique est maintenu et sert d'un guide visuel qui renforce la transmission du sens figuré.

Malgré l'utilisation de gestes, cette explication ne semble pas être efficace car Emilie utilise un vocabulaire trop complexe et introduit trop de nouveaux items lexicaux. Elle introduit « bourse » au milieu de son explication de l'expression, et introduit « petit boulot » tout de suite après. Elle accepte des réponses minimales des apprenants du type « oui » au lieu de vérifier si la compréhension est profonde, ce qui est hasardeux car les apprenants peuvent feindre la compréhension suite à une séquence de négociation même s'ils ne comprennent pas (Aston, 1986).

Par ces exemples, nous avons démontré que les gestes et mimiques faciales constituent une ressource pédagogique de grande valeur pour les apprentis-enseignants aussi bien que pour les apprenants lors des séquences d'explication lexicale. Ils servent à expliquer les sens concrets et figurés des items lexicaux, et de vérifier la compréhension d'items lexicaux inconnus. Nous faisons l'hypothèse que les gestes qui sont visibles sont conscientisés car ils sont mis dans le champ de la webcam. Les gestes observés ont tendance à être redondants à la parole orale, ce qui permet aux apprenants d'avoir un accès multimodal au sens.

[7] Là on assiste à une opportunité ratée pour encourager l'usage créatif (Joe, 1995, 1998). Emilie aurait pu dévoiler le sens littéral du verbe « arrondir » et faire deviner aux apprenants le sens figuré afin d'entraîner leur capacité à faire de l'usage créatif, c'est-à-dire accéder aux acceptions d'un item lexical qui ne sont pas apprises explicitement. La capacité à faire de l'usage créatif réduit le temps nécessaire pour apprendre du vocabulaire, car il est couteux en temps d'apprendre individuellement toutes les acceptions de chaque item lexical. Le développement de cette compétence devrait être visée et favorisée par tous les participants.

7.5 Conclusion

Nous nous sommes demandé comment les apprentis-enseignants se focalisent sur le sens des items lexicaux. Nous avons analysé à la fois les pratiques explicatives mises en place et l'usage des ressources sémiotiques. Les participants ont recours à l'anglais, le français, et l'usage d'images et de gestes pour expliquer le sens. Des moyens verbaux sont employés plus de trois fois plus souvent que des moyens visuels. L'approche multimodale nous a permis de démontrer que les quatre canaux offerts par le dispositif — le canal voco-verbal, le clavardage, la webcam, et l'envoi de documents — sont employés par les participants seuls ou en combinaison les uns avec les autres lorsqu'ils se focalisent sur le sens d'un item lexical. Or, ces trois derniers canaux occupent une place marginale dans certaines études qui portent sur l'explication (Fasel Lauzon, 2014 ; Miecznikowski, 2005). Par conséquent, une conscientisation de l'usage de ces ressources s'impose, aussi bien pour les chercheurs que pour les apprentis-enseignants. Nous avons remarqué, par la présence de commentaires méta-sémiotiques, un certain degré de conscientisation, mais il semble que cette conscientisation soit plutôt orientée vers les modes sémiotiques qui sont peu utilisés dans la vie quotidienne, à savoir l'envoi d'images. L'habitus didactique (Cicurel, 1993) ou le répertoire didactique (Cicurel, 2002) des apprentis-enseignants mérite d'être étendu pour incorporer une conscientisation de tous les modes, non pas seulement les modes novateurs ou peu utilisés comme l'envoi de documents.

Ce chapitre nous a permis d'aborder certaines particularités de la visioconférence. Les visages des apprentis-enseignants sont visibles en permanence, et c'est souvent la seule partie du corps qui est visible, et c'est la raison pour laquelle les expressions faciales sont amplifiées. Il incombe aux apprentis-enseignants de prendre conscience de l'effet amplificateur de la webcam, de même qu'il incombe aux utilisateurs de microphones de prendre conscience du fait que chaque bruit de bouche est capté et amplifié. Ces particularités peuvent être employées stratégiquement pour expliquer le sens d'un item lexical, comme le fait Samia dans l'exemple 22 en page 243 et comme le fait Adèle dans l'exemple 23 en page 247. Concernant les gestes, la position de repos est invisible, ce qui fait que la production de gestes demande une plus grande conscientisation. Quand les gestes sont produits dans le champ, ils ont tendance à être redondants à la parole, ce qui permet aux apprenants d'avoir un accès multimodal au sens. Enfin, l'envoi d'images demande non seulement un haut degré de conscientisation mais aussi une planification en amont de la séance. Pour cette raison, l'usage de cette affordance s'accompagne systématiquement de commentaires méta-sémiotiques.

Nous avons vu que l'utilisation d'une image ne dispense pas une explication en français, ce qui nous a amené à plaider pour le développement de la flexibilité multimodale. Cette compétence est à favoriser pour que des apprentis-enseignants puissent devenir des lexicographes multimodaux.

Chapitre 8 : Impliquer les apprenants

Dans les deux chapitres précédents, nous avons analysé la façon dont les apprentis-enseignants se focalisent sur la forme et sur le sens des items lexicaux qui sont deux facettes importantes de la connaissance d'un item lexical. Mais, comme nous l'avons évoqué dans le chapitre 2, l'interaction et la négociation sont des éléments cruciaux pour l'acquisition langagière. Les apprenants retiennent mieux les items lexicaux quant ils sont invités à participer activement à la séquence d'explication (Fasel Lauzon, 2014). La participation active des apprenants est d'autant plus importante que l'objectif principal de cette formation en ligne est l'entrainement à l'oral. Il incombe donc aux apprentis-enseignants de faire en sorte que les apprenants soient *impliqués* non seulement lors des activités communicatives, mais aussi lors des séquences de focalisation sur la langue comme les séquences d'explication lexicale.

L'implication des apprenants se joue aux niveaux cognitifs et socioaffectifs. Certains chercheurs ont tenté d'élaborer une grille pour mesurer l'implication des apprenants, notamment Laufer & Hulstijn (2001). Leur modèle a été repris et élaboré par Nation & Webb (2011). Selon ce modèle, *la charge d'implication* est fonction de trois éléments : le besoin, la recherche et l'évaluation (cf. chapitre 2, section 2.4.3.4 en page 61). A notre connaissance, ce modèle n'a pas été combiné avec une approche multimodale pour voir de quelle façon les ressources sémiotiques d'une plateforme de visioconférence sont employées pour impliquer les apprenants dans des séquences d'explication lexicale. En s'appuyant sur ce modèle, nous allons analyser la façon dont les apprentis-enseignants tirent profit des ressources de Visu pour impliquer les apprenants dans les séquences d'explication lexicale. La dimension socioaffective doit aussi être maintenue (Celik, 2008 ; Develotte & Mangenot, 2010; Mangenot, 2011). Guichon & Drissi (2008) soulignent la nécessité de « maintenir un équilibre socio-affectif et une fluidité dans les échanges ». L'apprenant doit se sentir à l'aise lors des séquences d'explication lexicale car la présence socio-affective est un prérequis à la présence cognitive (Jézégou, 2010).

Dans un premier temps, nous allons observer la façon dont les apprentis-enseignants impliquent les apprenants cognitivement en effectuant des rappels productifs et réceptifs. Dans un deuxième temps, nous analyserons des espaces de travail collaboratifs qui sont ouverts par les apprentis-enseignants suite à une panne lexicale ou à une erreur de production liée à un item lexical. Nous verrons la façon dont les apprenants sont impliqués pour

expliquer, trouver, évaluer et répéter un item lexical. Ensuite, nous analyserons la façon dont les apprentis-enseignants impliquent les apprenants en leur faisant réfléchir sur la culture ou sur la langue par des mots. Enfin, nous observerons la façon dont un apprenti-enseignant peut impliquer son apprenant en modifiant l'item lexical lui-même ou la définition de celui-ci. Tout au long de ces analyses, les dimensions multimodale et socioaffective seront prises en compte.

8.1 Rappels

Une façon d'impliquer l'apprenant cognitivement dans une séquence d'explication et de favoriser la mémorisation d'items lexicaux consiste à provoquer la recherche lexicale (Laufer & Hulstijn, 2001). Cette recherche a lieu quand la forme et le sens d'un item lexical ne sont pas donnés en même temps (Baddeley, 1993). Par exemple, on donne un mot à l'apprenant et on lui demande de trouver une définition. Inversement, on peut donner une définition à l'apprenant et lui demander de trouver un mot correspondant. Un rappel peut donc être réceptif (on donne le mot sans une définition et on va de la forme vers le sens) ou productif (on donne seulement une définition du mot et on va du sens vers la forme). Les rappels productifs sont plus couteux cognitivement.

8.1.1 Effectuer un rappel réceptif

Il y a rappel réceptif quand l'apprenti-enseignant donne la forme d'un l'item lexical mais non pas le sens. On trouve ce phénomène notamment dans des vérifications de compréhension qui peuvent être implicites ou explicites. Ces vérifications de compréhension constituent un bon moyen d'impliquer l'apprenant car elles témoignent d'un caractère proactif de l'apprenti-enseignant. Un tuteur réactif attend que l'apprenant pose des questions et réagit. Celik (2008 : 96) explique qu' « au contraire, un tuteur proactif ne se contente pas de répondre aux questions, mais doit solliciter les étudiants et s'appliquer à faire émerger des demandes ». Les rappels réceptifs abondent dans notre corpus et se trouvent dans quatre-vingt-onze séquences d'explication lexicale. Nous avons compté le nombre de séquences d'explication lexicale dans lesquelles chaque canal ou combinaison de canaux de Visu ont été utilisés pour initier un rappel réceptif d'un item lexical. Le plus souvent, le rappel réceptif est initié oralement par l'apprenti-enseignant, ou accompagné d'un rapprochement de l'écran, d'un haussement de sourcils, d'un visage d'incertitude ou d'un regard pensif. Moins souvent, le clavardage est utilisé pour écrire le mot afin que l'apprenant ait une représentation visuelle

pour faciliter la recherche lexicale. L'utilisation des canaux de *Visu* pour effectuer des rappels réceptifs se résume comme suit.

Canaux utilisés pour le rappel réceptif	**Nombre de séquences**
Microphone uniquement	46
Clavardage uniquement	1
Microphone + clavardage	6
Microphone + webcam	36
Microphone + clavardage + webcam	2
Total	**91**

Figure 1 : Canaux utilisés pour les rappels réceptifs

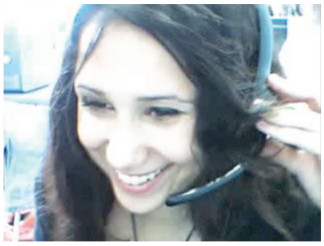
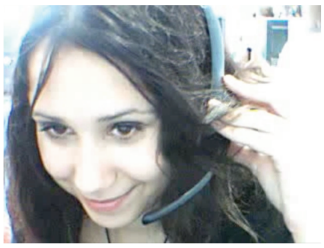
A deux occasions, trois canaux de Visu sont utilisés ensemble pour effectuer un rappel réceptif. Dans l'exemple 14 du chapitre 6 en page 193, Adèle utilise le microphone, la webcam et le clavardage pour vérifier la compréhension du mot « thym » quand elle explique sa manière de faire une boisson pour la toux à base de thym. En répétant le mot « thym » avec une intonation montante suivie d'une pause, en produisant une expression faciale d'incertitude, en inclinant la tête, et en écrivant un mot dont la compréhension est incertaine, Adèle implique Alannah dans l'explication. Elle utilise la multimodalité pour faire comprendre à l'apprenante qu'il s'agit d'un rappel réceptif. Alannah semble avoir compris, car elle produit le mot en anglais.

Les apprentis-enseignants peuvent aussi initier des rappels réceptifs par des vérifications de compréhension explicites (cf. chapitre 7, exemple 14 en page 233). Une vérification de compréhension peut également se faire de manière très discrète, en utilisant uniquement le clavardage, comme le fait Samia lorsqu'elle souligne par clavardage le mot « convaincre » afin d'en vérifier la compréhension (cf. chapitre 6, exemple 13 en page 192).

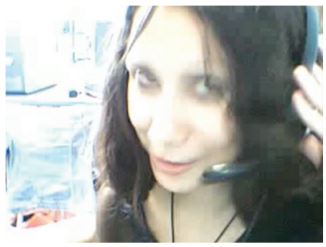
Quand on demande à un apprenant de tenter de faire un rappel réceptif et de donner une définition à un mot inconnu ou partiellement connu, l'apprenant peut être hésitant ou avoir peur de donner la mauvaise réponse[1], ce qui peut nuire à l'implication. Guichon (2017a : 57) plaide donc pour que l'apprenti-enseignant « [rende] visible la qualité de son écoute, sa disponibilité et sa patience pour encourager les étudiants à prendre part aux interactions orales avec confiance ». Observons la façon dont Emilie fait un véritable effort

[1] Par exemple, Aiden s'excuse quand il fait une erreur (voir exemple 6).

pour impliquer les apprenants dans une recherche lexicale quand elle vérifie la compréhension de l'expression « arrondir les fins de mois ».[2]

Transcription de la parole			Image du participant	Phénomènes observés
1		alors avant de passer à l'activité suivante/ (0,5) je vais vous donner une expression/ (0,5) et vous allez/ me dire si vous la connaissez/ ou pas\ (0,6) c'est une expression/ euh typiquement française/ (0,6) assez/ euh assez courante (.) donc vous l'entendrez (.) très très **très souvent** employée\ (0,9) arrondir/ les fins de mois\ (1,9)		Clin d'œil
2	FIN	euh (1,4)		
3	EMI	est-ce que qu'est-ce que ça peut [vouloi]r dire\ **(0,6)&**		Grand sourire
4	FIN	**[((rire))]**		
5	EMI	**&cherchez\ (1,9) ayez pas peur de vous: de vous tromper\ hein/ c'est pas grave/ (0,5) balancez n'importe quoi/ le le but c'est l'essayer de trouver\ (3,6)**		
6	FIN	**euh: (3,2)**		Sourire
7	ADN	**je pense que: (0,5) à la fin de: mois/ (0,3) euh: (0,3) chaque {persã} (0,3) ils sont: (0,8) or (0,5) ils recevont (.) euh ils recevont: (0,3) euh: (0,9) euh leur d'argent (1,0)**		

[2] La suite de cette séquence d'explication lexicale se trouve dans l'exemple 37 du chapitre 7.

8	EMI	leur argent/ oui leur salaire/ (0,6) leur [salai]re leur&		Regard caméra
9	ADN	[oui]		
10	EMI	&paye/ (.) oui/		
11	ADN	euh oui oui oui\		
12	EMI	c'est l'idée\ (0,4) c'est l'idée\ tu as compris l'idée\ (.) maintenant/ (.) arrondir les fins de mois (0,3) en tant que tel\ (0,7) qu'est-ce que ça peut vouloir dire\ (0,8) t'es étais sur la bonne voie c'est ça\ (0,4) fiona\ **(0,6)**(1,3)		
13	FIN	euh: (0,4)		
14	EMI	[°à ton avis\°]		
15	FIN	[je ne s]ais pas (.) hum (2,6) euh: qu'est-ce que c'est arrondir		

Exemple 1 : Une invitation à trouver le sens

Emilie commence par annoncer qu'elle va ménager un espace de travail sur le lexique : « je vais vous donner une expression, et vous allez me dire si vous la connaissez ou pas » (ligne 1). Ceci pousse les apprenants à écouter attentivement et leur permet d'anticiper l'activité mentale que requiert un rappel réceptif. Puis, avant même de donner la forme de l'item lexical, Emilie met en valeur l'utilité de celui-ci en disant que l'expression est « typiquement française », « assez courante », et « très très souvent employée » (ligne 1). Ceci encourage davantage les apprenants à écouter attentivement et avec anticipation. Ensuite, Emilie implique cognitivement les apprenants aux lignes 3 et 5 quand elle leur demande de chercher le sens de l'expression et d'émettre des hypothèses. Elle n'oublie pas le côté socioaffectif de sa demande quand elle les encourage avec un grand sourire en leur disant que ce n'est pas grave s'ils donnent une définition inexacte (ligne 5). Aiden émet une hypothèse (ligne 7) qui est ratifiée par Emilie avec encouragement et bienveillance (lignes 8, 10, 12). Emilie relance sa demande de rappel et invite Fiona à répondre avec un regard dirigé vers la caméra, ce qui donne à Fiona l'impression d'être regardée (ligne 12). Grâce à cette vérification de compréhension, Fiona demande le sens de l'item lexical (ligne 12). Le *besoin* est donc plus élevé car c'est l'apprenante qui en fait la demande.

La *charge d'implication* (Laufer à Hulstijn, 2001) est fortement élevée grâce aux stratégies mises en place par Emilie : il y a à la fois *recherche* (les apprenants cherchent le sens de l'item lexical), et *besoin* (Fiona demande une explication). Tout au long de la séquence, il y a un accompagnement socioaffectif de la part d'Emilie qui utilise sourires, regards caméra, et énoncés de rassurance (« ce n'est pas grave », « n'ayez pas peur »). Quand

Aiden émet une hypothèse, Emilie la répète pour valoriser sa contribution, ce qui est encourageant (Celik, 2008).

Si l'initiation d'un rappel réceptif est normalement une bonne manière d'impliquer l'apprenant, surtout si la dimension socioaffective est favorisée par des encouragements, nous avons repéré certaines tentatives de rappels réceptifs qui ont échouées. Afin qu'un rappel réceptif puisse avoir lieu, il appartient à l'apprenti-enseignant de laisser assez de temps à l'apprenant pour répondre à la vérification de compréhension, faute de quoi la séquence devient monogérée. Ce phénomène s'observe notamment chez Séverine, soit l'exemple suivant.

		Transcription de la parole
1	SEV	est ce que tu sais ce que c'est équipier (0,5) chez mcdonald's (0,4) équipier/ c'est les salariés\ (0,4) d'accord/ (0,6)
2	NAO	oui (0,8)
3	SEV	okay/ (0,3) alors\

Exemple 2 : Une vérification de compréhension avortée

Dans cet exemple, le rappel réceptif devient une vérification de compréhension fermée où l'apprenante n'a qu'à dire « oui » ou « non ». Il est impossible de déterminer si Naomi a exécuté mentalement un rappel réceptif avant que Séverine ait donné la réponse pendant le même tour de parole (ligne 1), mais cela semble peu probable.

Il faut laisser le temps de répondre à l'apprenant car en visioconférence, comme le rappelle Traverso (2011), les chevauchements sont plus fréquents en raison du décalage entre la production et la réception de la parole. La présence de chevauchements, ainsi que la difficulté que peuvent avoir les participants à gérer le passage du tour de parole, peuvent avoir une influence sur le mode qui est utilisé par l'apprenant pour participer à une séquence d'explication. Dans l'exemple suivant, Séverine laisse l'apprenante montrer qu'elle comprend l'item lexical en question (contrairement à l'exemple précédent), mais elle ne lui laisse pas assez de temps pour répondre oralement. Par conséquent, elle passe au mode écrit, ce qui va à l'encontre des objectifs pédagogiques du projet (s'entraîner à l'oral).

		Transcription de la parole
1	SEV	et agressif (.) est-ce que tu comprends (.) agressif\ (1,0)
2	NAO	°(inaud.)° (0,9)
3	SEV	euh
4	NAO	oh [oui oui oui] (0,3) [je] comprends\
5	SEV	[agress]i- (0,2) [oui] (0,3) qu'est-c`que ça veut dire agressif alors\ (0,4) dis-moi c`que [c'est]
6	NAO	[hu]m:

```
          (0,6)
7  SEV    e- (0,6) si toi tu es agressif/ comment tu parles\ (0,5) ((rire))
          (0,5)
8  NAO    euh:
          (0,4)
9  SEV    tu parles (1,4) tu es pas pol- (.) oui/ (.) vas-y (0,3) °euh° dis
          moi dans le chat °si tu°
          (0,5)
10 NAO    oh hum: (15,3)
          <((clavardage)) pour agressif je parlerais rude>
          (2,3)
11 SEV    oui/ (.) c'est ça\ (0,2) très bien\ (0,3) al[ors\]
12 NAO                                                [oui]
```

Exemple 3 : Une implication plus faible à cause des chevauchements

Séverine commence par vérifier la compréhension du mot « agressif », qui est l'un des mots-clés de l'activité[3] (ligne 1). Naomi dit qu'elle comprend le mot (ligne 4) et Séverine l'incite à le définir (ligne 5). Naomi se met à réfléchir (ligne 6), mais avant que celle-ci ne puisse répondre, Séverine reformule sa question et sollicite un exemple (ligne 7). Naomi se met à réfléchir à nouveau avec un regard pensif (ligne 8), mais Séverine lui demande aussitôt de répondre par écrit (ligne 9). On assiste à un épisode de ce que Sindoni (2013) nomme l'alternance des modes hétéro-déclenchée. Naomi prend le temps d'écrire sa réponse et donne un exemple qui témoigne de sa compréhension du mot (ligne 10). Séverine accepte la réponse écrite mais ne reprend pas la formule à l'oral pour valider la réponse (ligne 11). Ces deux derniers exemples montrent qu'impliquer l'apprenant nécessite aussi que l'on laisse le temps à celui-ci de répondre à une vérification de compréhension, faute de quoi l'implication cognitive de l'apprenant se dégrade (exemple 3) ou l'usage du mode oral est perdu (exemple 4).

Ces exemples corroborent le constat de Fasel Lauzon (2014 : 121) que « la grande majorité des demandes d'explication de vocabulaire des enseignants prend place aussitôt après l'introduction d'un lexème comme objet de discours dans l'activité en cours ». La compréhension des items lexicaux est donc une préoccupation première des apprentis-enseignants, et l'emploi de rappels réceptifs permet aux apprentis-enseignants d'impliquer leurs apprenants cognitivement.

8.1.2 Effectuer un rappel productif

Le rappel productif a lieu quand l'apprenti-enseignant donne une définition d'un mot (le sens) sans en donner la forme. Ceci implique les apprenants car ils doivent chercher la

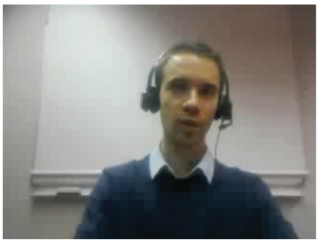
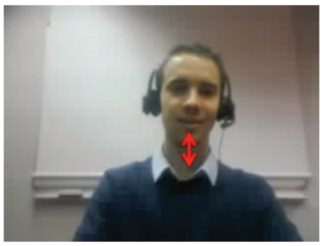
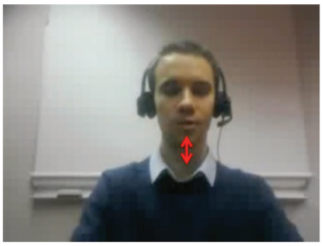
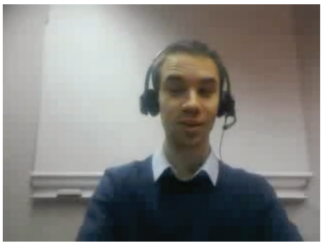
[3] Dans la troisième activité de la quatrième semaine, les apprentis-enseignants doivent réagir à différentes situations de conflit qui peuvent survenir en milieu professionnel. Les apprentis-enseignants jouent le rôle d'un manager et disent comment ils réagiraient face à des employés paresseux, agressifs, désagréables, ou en retard.

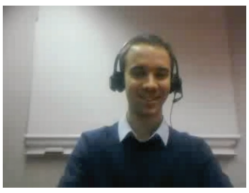
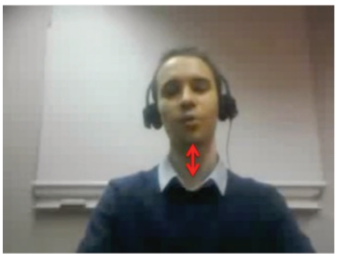
forme du mot (Laufer & Hulstijn, 2001). Nous avons repéré ce type d'implication dans quatre séquences d'explication lexicale, et le microphone est utilisé en combinaison avec la webcam à chaque fois.

Canaux utilisés pour le rappel productif	**Nombre de séquences**
Microphone + webcam	4
Total	**4**

Figure 2 : canaux utilisés pour les rappels productifs

Trois rappels productifs sont effectués par Mélissa. L'un se trouve dans l'exemple 12 ci-dessous quand elle recycle la définition d'Ana pour demander à Alejandra si elle connaît le mot. L'autre rappel productif est fait par Victor, quand il donne une définition des chaises musicales et demande à son apprenant s'il le connaît.

Transcription de la parole		**Image du participant**	**Phénomènes observés**
1	VIC	**est-ce que tu connais le jeu/ où hum tu sais il y a une musique (.) qui joue/ (0,6) et y a un nombre** limité de: de chaises/ (1,1)		Regard pensif
2	LIM	oui oui oui		
		(0,3)		
3	VIC	**d'accord**		
4	LIM	**euh:** (0,2) c'est le: même chose que (0,8) hum: (1,6) euh j- j'ai pensé sur le: même chose		Trois hochements de tête de tête et un sourire
		(0,3)		
5	VIC	**d'accord** (1,1) alors ça en français/ (.) ça s'appelle/ (1,3) la CHAISE MUSICALE\ (0,3)		Trois hochements de tête et regard vers le clavier
		enfin chez moi ça s'appelle comme ça <((clavardage)) La chaise musicale.> (0,6)		Haussement de sourcils

6 LIM 7 VIC	**((rire))** **((rire)) (1,6)**			Rires
8 LIM	c- et: comment vous dites en anglais/ **(1,6)** **hum: c'est: (0,3) "musical chairs"** **(2,4)**			Regard pensif, puis tête inclinée
9 VIC 10 LIM	**ah oui d'accord** (.) c'est la même chose\ (0,5) euh c'est la même chose oui			Hochement de tête et haussement de son épaule droite

Exemple 4 : Un rappel productif

Victor co-construit cette séquence d'explication avec Liam. Avec un regard pensif, Victor commence par donner une définition interrogative du jeu. Il invite au rappel et à la participation en parsemant sa définition de questions comme « tu connais ? » « tu sais ? » (ligne 1) et en attendant que l'apprenant réponde (ligne 2). Victor laisse Liam dévoiler l'état de sa compréhension et fait des hochements de tête pour en accuser la réception (lignes 4 et 5). Liam dit connaître le jeu, mais n'arrive pas à produire la forme. Victor la lui donne à oralement et par écrit (ligne 5). Liam rit ainsi que Victor. On assiste alors à ce que nous appelons un *alignement socioaffectif*. Pour terminer la séquence, Victor demande à Liam la traduction anglaise et incline la tête en attendant la réponse.

Dans cette séquence, Victor implique l'apprenant par différents moyens. Il sollicite d'abord un rappel productif en donnant le sens de l'item lexical « chaises musicales » sans en donner la forme. Ensuite il privilégie la production orale de l'apprenant en lui laissant le temps de répondre. On peut donc faire l'hypothèse que les alignements socioaffectifs (regards pensifs, rires communs) aident aussi à impliquer l'apprenant dans la séquence d'explication. Pour finir, Victor vérifie la compréhension en demandant une traduction, ce qui est un rappel réceptif. Victor implique Liam fortement en le mettant en position d'expert (Cappellini & Rivens Mompean, 2015) : c'est celui qui connaît le mot en anglais. D'ailleurs, dans sept séquences d'explication lexicale, l'apprenti-enseignant fait produire le mot en anglais, soit par

curiosité, soit pour vérifier sa propre compréhension avant de donner le mot juste. Quand un apprenti-enseignant demande le mot en anglais par curiosité, comme le fait Victor, on assiste à une réorganisation temporaire du contrat didactique, ce qui constitue un moyen d'impliquer l'apprenant.

Pour résumer, les rappels réceptifs sont actualisés par les apprentis-enseignants quand ils donnent un mot à leurs apprenants en leur demandant de trouver le sens. Ceci implique les apprenants cognitivement (Baddeley, 1993). Cette implication peut être mise en place par des vérifications de compréhension explicites ou implicites. La plupart sont faites en utilisant le microphone seul ou en combinaison avec la webcam. Certains rappels réceptifs sont actualisés en utilisant les trois canaux de Visu, comme nous avons vu chez Adèle quand elle vérifie la compréhension du mot « thym ». Il incombe aux apprentis-enseignants de fournir un accompagnement socioaffectif, comme le fait Emilie quand elle vérifie la compréhension de l'expression « arrondir les fins de mois ». Il revient aussi aux apprentis-enseignants de laisser assez de temps à leurs apprenants à la suite de la vérification, faute de quoi il n'y a pas d'implication (cf. exemple 3 en page 264). Les rappels productifs impliquent les apprenants cognitivement car ils doivent trouver la forme d'un item lexical à partir de son sens.

8.2 Espaces de travail collaboratif sur le lexique

Les vérifications de compréhension réceptives et productives arrivent souvent quand l'apprenti-enseignant estime qu'un mot est difficile à comprendre, ou quand l'apprenti-enseignant estime qu'un mot est bon à connaître. D'autres séquences d'explication lexicale émergent d'un besoin manifeste de l'apprenant, par exemple suite à une erreur de production ou une panne lexicale. En effet, les séquences de recherche lexicale ne se jouent pas uniquement dans la tête d'un interlocuteur, mais sont interactives (Goodwin & Goodwin, 1986). Quand une erreur survient, l'apprenti-enseignant peut la faire remarquer et ouvrir un espace de collaboration pour trouver un mot mieux approprié. A la suite d'une panne lexicale, l'apprenti-enseignant peut demander à l'apprenant de l'aider à trouver un mot juste. Il existe différentes façons d'impliquer les apprenants dans ces espaces collaboratifs. Les cinq sous-sections suivantes vont explorer ces différents types d'implication.

8.2.1 Faire remarquer un élément problématique

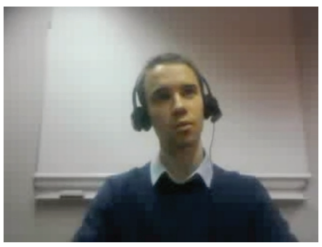
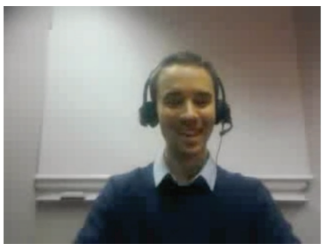
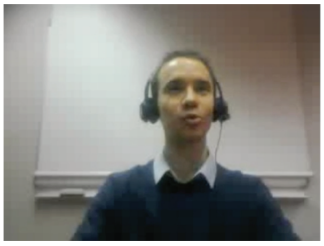
Une façon de provoquer une tentative d'auto correction est de faire remarquer un élément problématique dans le discours de l'apprenant. Nous avons repéré vingt-deux

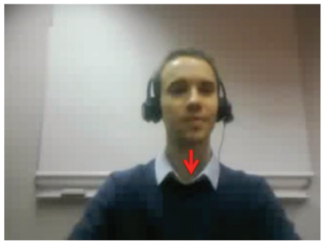
séquences d'explication lexicale dans le corpus où l'apprenti-enseignant fait remarquer un élément problématique. Le tableau ci-dessous indique les canaux utilisés.

Canaux utilisés pour faire remarquer un élément problématique	**Nombre de séquences**
Microphone uniquement	9
Microphone + webcam	13
Total	**22**

Figure 3 : Faire remarquer un élément problématique

La plupart du temps, le microphone est utilisé pour répéter le début de l'énoncé de l'apprenant jusqu'au problème, en combinaison d'un visage d'incertitude pour faire remarquer à l'apprenant qu'il y a un problème dans son *output*. L'exemple suivant est issu d'une discussion dans laquelle Victor et Liam réfléchissent sur la meilleure façon de réprimander un employé qui arrive systématiquement en retard.

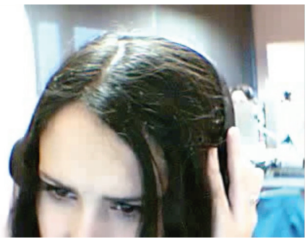
Transcription de la parole	**Image du participant**	**Phénomènes observés**
1 LIM euh: je te do- euh: (1,1) je: le donne (1,2) euh: (.) une (.) caution/ (1,9) 2 VIC **euh: non/**		Regard pensif
((rire)) (0,4) [ça c'est le mot en] anglais 3 LIM **[((rire))] (0,8)** 4 VIC **[hum:]** 5 LIM [euh: a-] (0,5) atten- attention/ (0,8)		Rires
6 VIC **euh une euh un** AVERTISSEMENT (1,2) 7 LIM avertissement/ (0,4) 8 VIC oui <((clavardage)) Un avertissement.> (1,6)[donner/ (.) un aver]tissement\ 9 LIM [euh: oui] (0,9) je le donne (0,6) un: (0,8) avertissement\ (0,5)		Regard pensif

10 VIC **très bien**		Hochement de tête

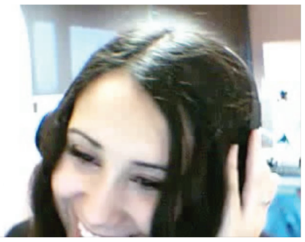
Exemple 5 : Pointage d'une erreur

Liam emploie le mot « caution », qui n'a pas du tout le même sens que le mot en anglais[4] (ligne 1). Victor, avec un regard pensif, fait remarquer à Liam que ce mot est inexact (ligne 2). En riant, il explique que ce n'est pas la même chose qu'en anglais. Liam s'implique et tente alors une auto correction (ligne 5). Victor donne oralement le mot juste (ligne 6) et par écrit (ligne 8), ainsi que la collocation « donner un avertissement » (ligne 8). Liam répète la forme correcte (ligne 9) et se fait féliciter par Victor (ligne 10). Le fait de souligner une erreur, comme le fait Victor, sert à ouvrir un espace de collaboration et suffit pour impliquer l'apprenant.

Les apprenants peuvent ressentir une gêne et un embarras quand l'apprenti-enseignant dirige l'attention vers une erreur comme un calque langagier. Pour pallier cela, un accompagnement socioaffectif s'impose. Emilie a l'habitude d'employer des moyens verbaux, utilisant le microphone plutôt que la webcam pour l'accompagnement socioaffectif. A plusieurs reprises, elle encourage oralement ses apprenants lors des séquences d'explication lexicale, comme on l'a vu dans l'exemple 1 ci-dessus. Dans l'exemple suivant, Aiden est en train de parler des progrès langagiers qu'il fait grâce aux bilans envoyés par Emilie, quand Emilie souligne un élément problématique dans son discours.

Transcription de la parole	**Image du participant**	**Phénomènes observés**
1 ADN et j'ai laissé (0,2) euh (0,7) très très plus (0,3) et: (4,5) 2 EMI tu l'as **fait très très/** (1,6) [euh]: 3 ADN [très très] (3,0) très {pluz}: (0,3) informations (1,2)		Froncement de sourcils

[4] *Caution* en anglais veut dire « prudence » ou « circonspection ».

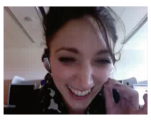
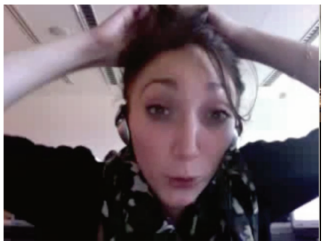
4	EMI	AH d'accord\ avec BEAUCOUP PLUS d'informa[tions\] ((rire))&		Sourires
5	ADN	[mh\]		
6	EMI	&d'accord\ très très plus-/ c'est très anglais\ ((rire))		
7	ADN	mh\ oui oui/ ((rire)) (0,2)		
8	EMI	mais c'est pas la même [chose que le "very mu]ch"\&		
9	ADN	[désolé\] (0,5) [d- dé]solé\&		
10	EMI	&[((rire))]		
11	ADN	&((rire)) (0,7)		
12	EMI	{nã} ne t'excuse pas/ c'est tout à fait normal\ (0,7) dis-toi bien que quand moi je parle anglais je fais la même chose [hein ((rire))]		
13	ADN	[mh mh mh] mh\ (1,1)		
14	EMI	c'est normal\ (0,6) bon\ alors on en reparlera		

Exemple 6 : Encouragement oral

Emilie fait remarquer un élément problématique dans le discours de l'apprenant en répétant le début de l'énoncé de celui-ci jusqu'au problème avec une intonation montante (ligne 2). Son visage n'est pas très visible, mais on constate un froncement de sourcils. Puis, en souriant, Emilie remarque que cela vient de l'anglais, faisant l'hypothèse qu'Aiden a fait un calque dans sa tête. Aiden s'excuse, et Emilie l'encourage en disant que c'est normal et qu'elle rencontre les mêmes difficultés quand elle s'exprime dans la langue maternelle de l'apprenant (ligne 12). Emilie se met ainsi au même niveau que l'apprenant, ôtant un déséquilibre qui se met en place souvent entre enseignant et apprenant dans certaines contrats didactiques classiques. Le contrat didactique qui se met en place ici, en visioconférence, fait en sorte que les apprentis-enseignants encouragent les apprenants et se mettent à leur niveau par des implications socioaffectives.

L'accompagnement socioaffectif suite au repérage d'un élément problématique dans le discours de l'apprenant peut aussi se faire par des rires et sourires. Dans l'exemple suivant, Adèle implique Alannah de plusieurs façons. La séquence qui suit vient au début de l'enregistrement de la deuxième semaine. Ne disposant pas des tours de paroles précédents[5], nous supposons qu'Adèle avait demandé à Alannah ce qu'elle avait fait pendant le week-end.

[5] Le début de la séquence n'a pas été enregistré.

Transcription de la parole	Image du participant	Phénomènes observés
1 ALN (c'est) le week-end/ (.) donc euh c'était un peu xxx (1,1) 2 ADL un peu comment/ **(2,1)** 3 ALN **euh (0,3) ((rires))**		Rapprochement de l'écran, yeux plissés, tête tournée et tête inclinée
4 ADL **[((rires))]** 5 ALN **[en (.) ((rires))] <((en riant)) "I can't even (say; think)">**	 	Rires
6 ADL **ENNUYEUX ENNUYEUX** 7 ALN **(ennuyeux) (.) c'est ça**		Rapprochement de l'écran et haussement de sourcils
8 ADL [ah:\] 9 ALN [°(oui)°] (0,3) ((rires)) 10 ADL °tu° (0,3) redis-le (1,0) EN-NUY-EUX (0,4) 11 ALN en- (.) -nuy- (.) -eux (.) ((rires)) (0,2) 12 ADL **en-/**		Prononciation exagérée avec la bouche ouverte
(0,5) 13 ALN ((rires)) 14 ADL -nuy-/ **(0,3)** 15 ADL **((rires))** (0,5) ((rires))		Tête inclinée après la deuxième syllabe

Exemple 7 : Un partage d'émotions

Alannah prononce de façon incompréhensible le mot « ennuyeux » (ligne 1), ce qui amène Adèle à demander une répétition en s'approchant de l'écran et en produisant une expression faciale qu'elle a l'habitude de faire quand elle demande une répétition (ligne 2). Alannah se met à rire, et les deux rient ensemble (lignes 3-5). Adèle donne la forme correcte et sollicite une répétition d'Alannah, syllabe par syllabe (lignes 6-10). Alannah répète le mot, syllabe par syllabe (ligne 11).

Cet exemple montre comment la webcam peut être utilisée en combinaison avec le microphone pour alerter l'apprenant sur une erreur de production. En demandant

explicitement une répétition, ou en répétant l'énoncé de l'apprenant jusqu'à l'élément problématique, Adèle a l'habitude de se rapprocher à l'écran, de tourner sa tête et de produire des expressions faciales semblables à celles que l'on observe aux lignes 2-3 dans l'exemple ci-dessus. Cette combinaison de ressources est aussi employée souvent par Mélissa et par Victor pour diriger l'attention de l'apprenant vers un élément problématique.

On observe également dans cet exemple un accompagnement socioaffectif. Quand Alannah commence à rire, Adèle se met aussi à rire (ligne 4). Adèle mime ainsi les émotions d'Alannah pour qu'elle ne soit pas la seule à rire. Les émotions sont véhiculées par les sourires et les rires, et pendant ces quelques secondes, la webcam et le microphone sont dédiés uniquement au partage d'émotions. Adèle garde le sourire quand elle demande à Alannah de répéter la forme correcte du mot « ennuyeux ». Certains apprentis-enseignants comme Adèle, Mélissa et Victor utilisent la webcam pour réaliser cette implication socioaffective, en produisant sourires, rires, et rapprochements à l'écran. D'autres comme Emilie utilisent le microphone pour encourager. Nous disons avec Celik (2008) que ces adoucisseurs et encouragements servent à impliquer et à motiver les apprenants, et à minimiser la gêne occasionnée par le fait de se faire corriger.

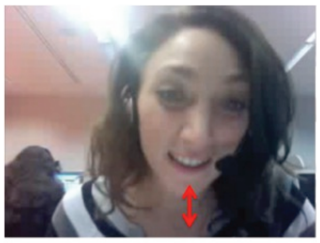
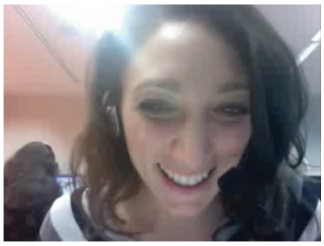
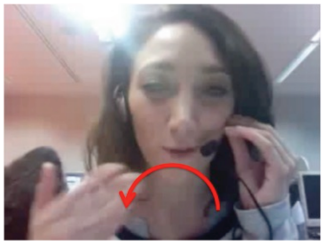
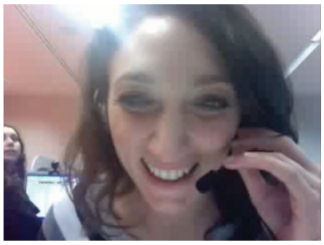
8.2.2 Faire expliquer à l'apprenti-enseignant

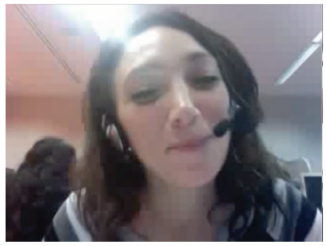
Quand un apprenant rencontre une panne lexicale, l'apprenti-enseignant peut faire en sorte que l'apprenant joue un rôle actif dans la résolution de celle-ci. Notamment, l'apprenti-enseignant peut demander à l'apprenant d'expliquer le mot recherché, ce qui oblige les apprenants à s'impliquer dans la séquence en tant qu'explicateur et à produire de l'output langagier. Celui-ci a plusieurs avantages, notamment le développement de l'aisance à l'oral, le repérage de lacunes langagières, et la vérification d'hypothèses (Swain, 1995). Nous avons repéré ce phénomène dans vingt-huit séquences d'explication lexicale, et l'utilisation des canaux de Visu se résume comme suit.

Canaux utilisés pour faire remarquer un élément problématique	**Nombre de séquences**
Microphone uniquement	17
Microphone + webcam	11
Total	**28**

Figure 4 : Faire expliquer à l'apprenti-enseignant

La plupart des demandes d'explication se font uniquement à l'oral, mais sont parfois accompagnés d'expressions faciales, rapprochements de l'écran, ou gestes, comme dans l'exemple suivant, où Catriona ne connaît pas le nom du métier de son père. Adèle lui demande alors de l'expliquer.

Transcription de la parole	**Image d'Adèle**	**Phénomènes observés**
1 CTR et mon père il euh (0,2) il travaille dans l'école mais avec le (0,4) le travaux **(oui oui)** (0,6) je ne sais pas le mot en français/ (.) mais hum\		Hochement de tête
2 ADL alors\ (0,4) **décris-moi/ et je vais t'aider à trouver le mot\ (1,1)** 3 CTR **hum** (5,0)		Rapprochement de l'écran et sourire
4 ADL ou décris en parlant (0,5) **en parlant\ (0,5)**		Geste circulaire représentant l'acte de parler
5 CTR "oh ok" (0,4) **hum: il (0,5)** il fait des répa- les les répares dans l'école **hum si: (0,8) si: quelque chose est cassée: (.) il le (0,3)**		Rapprochement de l'écran et sourire
6 ADL OUI:\ (.) **[euh:]** 7 CTR **[il le répare:]** (0,3) 8 ADL **alors euh peut-être on appellerait ça:/** hu:m\ (1,0) agent/ de maintenance\ (1,1) 9 CTR ah oui de maintenance oui\ (0,3)		Regard pensif

10 ADL	<((clavardage)) agent de maintenance> un (0,6)		Sourire
11 CTR	oui		
12 ADL	agent de maintenance\ (0,3) **[((rires))]**		
13 CTR	**[oui mais il travaille] dans un école (0,2) oui (0,5)**		
14 ADL	**très bien/** (0,3) donc euh tes deux parents sont à l'école\		

Exemple 8 : Demande d'explication

Avant même la panne lexicale, Adèle pousse son apprenante à parler en faisant des hochements de tête servant de régulateurs d'écoute (ligne 1). Quand Catriona commente la panne lexicale qu'elle rencontre, Adèle lui demande de décrire le mot recherché (ligne 2), ce qui la met en position d'explicatrice. A la suite de sa demande, Adèle sourit et se rapproche à l'écran, ce qui implique Catriona en lui donnant l'impression qu'elle est écoutée (ligne 3). Il s'ensuit alors une longue période de silence pendant laquelle Catriona regarde son clavier. Puis Adèle l'invite à reprendre la parole. Cette consigne méta-explicative et méta-sémiotique se réalise à l'aide d'un geste (ligne 4). Quand Catriona décrit le travail de son père, Adèle se penche à nouveau vers l'écran et sourit (ligne 5). Après l'explication de Catriona, Adèle réfléchit pensivement (lignes 6-8) et donne oralement le mot juste et par écrit (lignes 8, 10, 12). Elle utilise l'adoucisseur « peut-être » (ligne 8) quand elle donne le mot au lieu de le donner avec un ton péremptoire. Adèle termine la séquence par un commentaire qui témoigne de l'attention qu'elle porte à la parole de Catriona.

L'implication effectuée par Adèle se fait à trois niveaux. Au niveau cognitif, Adèle implique son apprenante en la mettant en position d'explicatrice. Elle ouvre un espace de collaboration dans lequel l'apprenante peut expliquer le mot recherché. Cette ouverture se fait oralement (« décris-moi », ligne 2) et visuellement, avec un rapprochement à l'écran et un sourire. Le microphone et la webcam sont utilisés ici de manière complémentaire. Puis, au niveau sémiotique, Adèle implique l'apprenante en lui demandant d'utiliser le microphone plutôt que le clavardage pour s'exprimer (« décris en parlant », ligne 4). Cette demande se fait à l'aide d'un geste. Là aussi, deux modes sémiotiques sont combinés pour réaliser une action. L'implication permet à l'apprenante à la fois de s'entraîner oralement, en expliquant le travail que fait son père, et de trouver le mot juste recherché. Enfin, au niveau socioaffectif, Adèle encourage son apprenante en se rapproche à l'écran en souriant quand cette dernière explique ce que fait son père.

8.2.3 Faire évaluer une définition ou une proposition d'un mot juste

A la suite d'une panne lexicale ou d'un élément problématique dans le discours de l'apprenant, un travail collaboratif s'impose pour trouver un mot plus approprié. Dans la section précédente, nous avons vu comment l'apprenti-enseignant peut impliquer son apprenant en lui faisant expliquer le sens du mot recherché. Une fois le mot trouvé, l'apprenti-enseignant peut impliquer l'apprenant davantage en lui faisant évaluer une proposition d'un mot juste ou d'une définition. Dans le chapitre précédent (exemple 35), nous avons vu un exemple où Adèle vérifie, par des gestes, sa propre compréhension de ce qu'Alannah cherche à dire. Elle donne une définition de l'item lexical « les attractions » en décrivant, à l'aide de gestes, une aire de jeux pour enfants. Ainsi, elle fait évaluer à son apprenante le sens de l'item lexical « les attractions » avant de donner la forme de celui-ci. Nous avons observé ce même phénomène dans l'exemple 25 du chapitre précédent en page 251 où Mélissa fait évaluer, par des gestes, une définition du verbe « bâiller » à son apprenante avant de donner la forme de celui-ci. Ce phénomène a été repéré dans sept séquences d'explication lexicale. Il est intéressant de constater que les gestes sont employés par les apprentis-enseignants cinq fois sur sept lors qu'ils font évaluer une définition à leurs apprenants. Dans la sixième séquence où la webcam est utilisée, Mélissa regarde directement la caméra quand elle fait évaluer une définition de la pause café. Le tableau ci-dessous résume l'utilisation des canaux pour ce type d'implication.

Canaux utilisés pour	**Nombre de séquences**
Microphone uniquement	1
Microphone + webcam	6
Total	**7**

Figure 5 : Faire évaluer le sens d'un mot

La webcam est donc fortement utilisée pour faire évaluer aux apprenants le sens d'items lexicaux. Quand les apprentis-enseignants savent qu'il est fort probable que l'apprenant ne connaisse pas le mot en question, ils ont tendance à utiliser des gestes pour faire passer le sens.

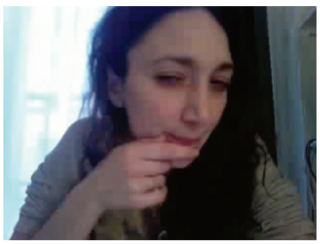
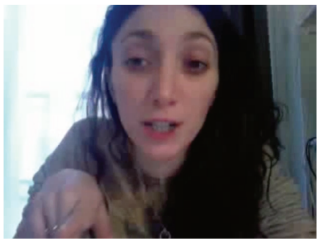
Cette implication marche le plus souvent dans l'autre sans, lors des pannes lexicales, où l'apprenti-enseignant fait des propositions de mots justes. L'apprenant est impliqué cognitivement quand il doit évaluer l'acceptabilité sémantique et syntaxique d'un mot dans un contexte donné (Laufer & Hulstijn, 2001). Nous avons repéré ce phénomène dans cinquante-

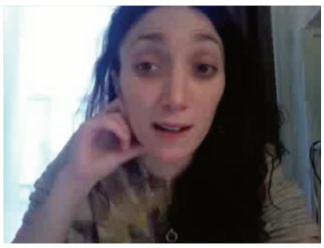
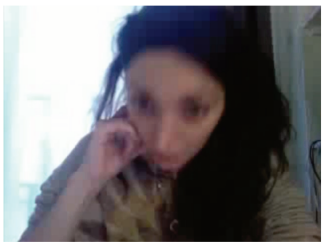
cinq séquences d'explication lexicale. Le plus souvent, cette stratégie d'implication est faite uniquement par le microphone. Quand la webcam est utilisé, c'est souvent pour donner un regard direct vers la caméra, faire des haussements de sourcils ou produire un visage d'incertitude. Le clavardage est utilisé huit fois sur cinquante-cinq pour poser par écrit des propositions de mots que l'apprenant doit évaluer.

Canaux utilisés pour	**Nombre de séquences**
Microphone uniquement	20
Clavardage uniquement	4
Microphone + clavardage	8
Microphone + webcam	13
Microphone + clavardage + webcam	10
Total	**55**

Figure 6 : Faire évaluer une proposition de mot juste

Observons dans un premier temps la façon dont Adèle utilise sa voix, sa main et son visage pour impliquer son apprenante dans la résolution d'une panne lexicale. L'exemple suivant est issu de la cinquième semaine où les apprentis-enseignants doivent réfléchir à la création d'une entreprise fictive, le camion-restaurant. Alannah cherche à dire « marché-cible » ou « clientèle » mais dit « cible marché », ce qui s'apparente à un calque syntaxique de sa traduction anglaise (*target market*). Adèle implique l'apprenante en la mettant en position d'évaluatrice de la manière suivante :

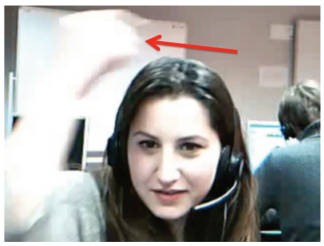
Transcription de la parole	**Image du participant**	**Phénomènes observés**
1 ALN pour notre euh: "like" (0,6) cible marché/ (1,8) 2 ADL **mh/** 3 ALN **cible/** **(1,4)**		Adèle se rapproche à l'écran, touche son menton et produit une expression faciale d'incertitude.
4 ADL **supermarché** (1,0) nan/ 5 ALN non non non\ (0,5) c'est (1,0) <((clavardage)) cible marche?> (1,2)		Geste déictique vers l'écran

6	ADL	cible marché\ (0,6) ah\ (.) donc la cible marché ça va être (.) **la clientèle/ (1,2)**		Haussement de sourcils
7	ALN	"yeah target market" **(0,3)** "**so**"		Hochement de tête

Exemple 9 : Faire évaluer des mots justes

A la suite de l'énoncé qu'Adèle ne comprend pas, celle-ci se rapproche à l'écran, se tourne la tête, plisse les yeux et touche son menton (ligne 2) afin d'ouvrir un espace de collaboration dans lequel la séquence de négociation peut avoir lieu. Alannah est impliquée dans un premier temps car elle doit trouver un moyen de s'exprimer autrement. Adèle propose ensuite le mot juste « supermarché » qui ressemble phonologiquement à la façon dont Alannah a prononcé « cible marché » (ligne 1). Sur ce, Adèle fait un geste déictique vers l'écran (ligne 4). Ce geste aide à communiquer et à rendre saillant le fait que ce soit une proposition qui doit être évaluée par l'apprenante. Alannah refuse cette proposition de mot juste et écrit dans la fenêtre de clavardage « cible marché ». Puis, Adèle vérifie sa propre compréhension du mot recherché et fait évaluer un nouveau mot à l'apprenante : le mot « clientèle » (ligne 6). Adèle hausse les sourcils et utilise une intonation montante, ce qui oblige l'apprenante à évaluer l'acceptabilité du mot. L'intonation de la voix, les mouvements de sourcils, les gestes, et la proxémique constituent autant de moyens par lesquels Adèle implique l'apprenante en lui demandant d'évaluer des propositions de mots justes.

Parfois, l'apprenti-enseignant fait évaluer à la fois la forme et le sens d'un item lexical, soit l'exemple suivant, où Angela cherche le mot « couronne ».

Transcription de la parole			Image du participant	Phénomènes observés
1	ANG	des: petits gâteaux et une hum: (2,2) **(inaud.)** (.) pour le: (0,2)		Angela tend l'index et maintient la main au-dessus de la tête
2	SEN	chapeau/		
3	SAM	un c- (.)un clown/ (0,2)		
4	SEN	x		
5	SAM	ah\ chapeau/		
6	ANG	non pas un chapeau un: (0,7) euh: (4,5) une c- (0,5) une couronne/ (0,7) c'est ça/ (1,5)		Samia tend l'index et dessine un chapeau au-dessus de la tête
7	SAM	une couronne/ **(1,1)**		
8	ANG	oui (0,2)		
9	SEN	°couronne:/°		
10	SAM	oui/ (0,2) c'est ça une couronne\		

Exemple 10 : Faire évaluer la forme et le sens d'un item lexical

Angela cherche le mot « couronne » lorsqu'elle propose des activités pour un goûter d'anniversaire. Suite à une pause longue de deux secondes, elle fait un geste iconique qui représente le mot recherché (ligne 1). Son binôme Sean propose le mot « chapeau » (ligne 2), et Samia propose le mot « clown » (ligne 3). Samia répète la proposition de Sean avec une intonation montante (ligne 5). Angela doit donc évaluer ces propositions. Angela refuse la proposition du mot « chapeau » (ligne 6) et cherche sur Internet pendant plus de quatre secondes avant de trouver le mot « couronne ». Elle le prononce de manière incertaine avec une intonation montante. Samia fait évaluer le mot « couronne » avec une intonation montante, et fait évaluer son sens par un geste iconique semblable à celui fait par Angela (ligne 7). Angela accepte la proposition (ligne 9) et Samia la ratifie (ligne 10). Cet exemple montre qu'on peut impliquer l'apprenant en lui demandant d'évaluer la forme ou le sens d'un item lexical, et que la webcam est un bon outil quand on demande à un apprenant d'évaluer le sens.

8.2.4 Faire répéter la forme correcte

Quand le mot juste est trouvé, l'implication de l'apprenant n'est pas terminée. Dans quinze séquences d'explication lexicale, l'apprenti-enseignant demande à l'apprenant de répéter la forme correcte, ce qui favorise la mémorisation. En effet, répéter un item lexical renforce la connaissance productive de celui-ci, ce qui constitue un type de connaissance plus robuste que la connaissance réceptive (Webb, 2005, 2009). Nous avons observé ce

phénomène dans l'exemple 7 ci-dessus. L'exemple le plus riche de ce phénomène a été présenté dans le chapitre 6 (exemple 22) où Adèle demande à Alannah de répéter le mot « prononciation » syllabe par syllabe en utilisant son visage, ses mains, sa voix et le clavardage. Le microphone, la webcam et le clavardage sont utilisés ensemble pour faire répéter l'item lexical quatre fois sur quinze. La plupart du temps, la webcam est utilisée pour faire répéter, en faisant des mouvements de bouche exagérés, en se rapprochant à l'écran, et en haussant les sourcils.

Canaux utilisés pour	**Nombre de séquences**
Microphone uniquement	5
Microphone + webcam	6
Microphone + clavardage + webcam	4
Total	**15**

Figure 7 : Faire répéter oralement l'item lexical

Pour résumer, les exemples précédents montrent comment les apprentis-enseignants, suite à une panne lexicale ou à une erreur de production liée au lexique, ouvrent un espace de collaboration dans lequel ils impliquent les apprenants en leur faisant remarquer l'élément problématique, en leur faisant expliquer le mot recherché, en leur faisant évaluer la forme ou le sens d'un mot, et en leur faisant répéter la forme correcte de celui-ci. Parfois, un accompagnement socioaffectif s'observe, ce qui réduit la gêne occasionnée par le fait de se faire corriger. Nous allons voir dans la section suivante que les apprentis-enseignants peuvent aussi ouvrir des espaces de collaboration entre apprenants.

8.2.5 Faire expliquer à l'autre apprenant

Un apprenant peut-il progresser dans une langue étrangère en dialoguant avec un autre apprenant ? Les avis sont partagés. Porquier & Py (2008) soutiennent que l'interaction entre deux apprenants ne suffit pas pour progresser, car il n'y a pas de tension entre l'interlangue[6] de l'apprenant et la compétence native souhaitée qu'aurait un enseignant. De manière générale, le travail collaboratif entre apprenants a l'avantage d'augmenter le temps de parole de chacun (Mangenot, 2011). Fasel Lauzon (2014) a trouvé que les noyaux polygérés sont rares mais intéressants d'un point de vue interactionnel. La collaboration entre apprenants facilite l'apprentissage pour plusieurs raisons. Kim (2008) souligne que la collaboration permet la focalisation sur la forme et a démontré expérimentalement que les tâches

[6] L'interlangue est la « compétence en voie d'élaboration » (Porquier et Py, 2008 : 28).

collaboratives engendraient plus d'acquisition lexicale que les tâches individuelles. Selon cette auteure, l'input d'un apprenant lors d'une tâche collaborative peut être aussi efficace que l'input d'un enseignant pour l'acquisition langagière. Comme le rappelle Newton (2001), la collaboration entre apprenants est bénéfique car l'étendue lexicale d'un groupe est toujours plus large que celle d'un individu. Enfin, en cas de détresse lexicale, l'apprenant le plus fort peut aider l'apprenant le plus faible (Folse, 2004).

Le travail collaboratif entre apprenants peut être facilité par la visioconférence. Fasel Lauzon (2014) mentionne que dans une salle de classe, il peut être difficile pour un apprenant de donner son explication à un autre apprenant et non pas à l'enseignant. En effet, la disposition de la salle de classe en rangs ne permet pas aux apprenants de se voir et de communiquer. Dans notre corpus, les participants qui interagissent en binôme ou en trinôme voient tous les interlocuteurs en permanence, ce qui constitue un avantage de la visioconférence.

Samia semble être consciente de l'aide qu'un apprenant peut apporter à l'autre, comme en témoigne la consigne générale suivante que l'on retrouve pendant l'interaction de la quatrième semaine.

		Transcription de la parole
1	SAM	euh angela je vais te demander euh: (0,8) d'essayer je suis désolée (0,3) angela je vais te demander/ (0,5) de: de mh: (0,6) d'aider en fait sean (0,4) euh: quand^euh quand tu vois qu'il est un peu en difficulté d'accord/ (1,1) donc si [tu pouvais refor-]
2	ANG	[d'accord je] vais essayer (0,5)
3	SAM	si tu pouvais reformuler euh (0,4) quand il comprend pas ça ça: ça pou- (0,4) ça ça pourrait être utile

Exemple 11 : Sollicitation de l'aide de l'apprenante la plus forte

Samia semble être gênée de déléguer son rôle d'enseignante, mais elle estime qu'Angela serait parfois plus capable d'aider Sean.

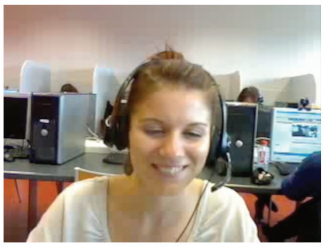
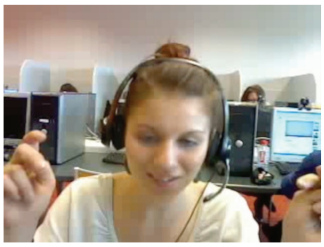
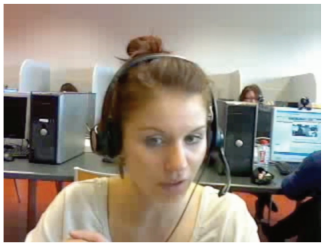
On retrouve par ailleurs des sollicitations ponctuelles de la part des apprentis-enseignants : dans notre corpus, on retrouve la collaboration entre apprenants provoquée par l'apprenti-enseignant dans sept séquences d'explication lexicale.

Canaux utilisés pour	Nombre de séquences
Microphone uniquement	5
Microphone + webcam	2
Microphone + clavardage + webcam	1
Total	**8**

Figure 8 : Faire expliquer à son binôme

La plupart de ces demandes d'explication se font uniquement par le microphone. Mélissa regarde directement vers la webcam trois fois quand elle demande à une apprenante d'aider son binôme, et utilise le clavardage une fois pour donner un support visuel à l'explicatrice.

Lors des interactions en trinômes, de longs dialogues entre l'apprenti-enseignant et un des apprenants surviennent souvent, ce qui entraîne l'autre apprenant à être spectateur et non pas participant actif. Quand une recherche lexicale doit être faite, on peut impliquer l'apprenant qui était jusqu'alors spectateur comme une aide pour trouver le mot juste. C'est ce que fait Mélissa dans l'exemple suivant, quand Ana cherche le mot « collègue » mais ne le trouve pas.

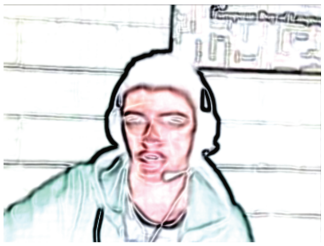
Transcription de la parole			Image du participant	Phénomènes observés
1	ANA	amé- euh améliorer notre: compétence/ avec notre: (0,2) camarade de: travail/ (.) non c'est pas ca- (.) camar[ade c'est]: (0,5) [hum:]		Yeux plissés et grand sourire
2	MEL	[nan/] **(0,8) [alejandra est-ce que t]u connais le mot/ (0,2)**		
		les (.) cama:**RADES (.) de travail/** comment on appelle ces personnes/ (0,3) les		Geste emblématique lors du discours rapporté
		[gens qui **travaillent avec**]&		Tête tournée et inclinée vers l'écran
3	ALJ	**[euh:]**		
4	MEL	**&toi/** (5,9) j'vais vous l'écrire c'est le mot/ (.) collègue\ (0,5)		
5	ANA	euh [les collègues]		
6	ALJ	[ah: collègue\]		
7	MEL	<((clavardage)) collègue> (0,5)		

8	ANA	euh		
9	MEL	un collègue/ et au pluriel (.) des/ collègues\		

Exemple 12 : Une sollicitation de l'aide d'une apprenante pour donner la forme

Mélissa utilise sa voix, ses gestes et sa tête pour impliquer Alejandra, qui était jusqu'alors spectatrice, dans la séquence. Ana cherche à dire le mot « collègue », mais ne trouve pas la bonne formulation (ligne 1). Elle dit « camarade de travail » au lieu de « collègue », faisant ainsi ce que Lüdi (1994) nomme le *bricolage lexical*. Mélissa sourit quand elle se rend compte de ce qu'Ana cherche à dire et de l'opportunité pédagogique qui se présente ainsi (ligne 2). Puis, Mélissa cite Ana en faisant des guillemets gestuels pour représenter le discours rapporté et le bricolage lexical de celle-ci. En citant Ana, Mélissa valorise son travail fourni, ce qui constitue un moyen de soutenir la dimension socioaffective de l'échange (Celik, 2008). Mélissa donne une définition d'un collègue : « les gens qui travaillent avec toi » et penche sa tête vers la webcam pour solliciter une réponse d'Alejandra (lignes 2-4). En déployant ces ressources sémiotiques, Mélissa provoque le rappel productif d'Alejandra. Quand aucune réponse n'est donnée pendant la pause de six secondes, Mélissa donne le mot juste à l'oral (ligne 5) en le répétant par écrit (ligne 7). Enfin, elle aborde le fonctionnement grammatical du mot en le mettant au pluriel (ligne 9). Mélissa joue le rôle d'arbitre pour faire collaborer les apprenantes entre elles, même si c'est finalement Mélissa qui donne le mot.

Dans l'exemple précédent, Mélissa sollicite l'aide d'une apprenante pour le rappel productif, du sens vers la forme. Un apprenti-enseignant peut également solliciter l'aide d'un apprenant pour le rappel réceptif, de la forme vers le sens, comme le fait Étienne dans l'exemple suivant. Lors de la quatrième semaine des interactions, les apprenants pensent à des activités à mettre en place pour un goûter d'anniversaire. Sophie propose que les enfants jouent à cache-cache, ce qui est accepté par Étienne car ils ont un budget limité et ce jeu n'est pas coûteux à mettre en place. Immédiatement après, Conor, qui était jusqu'alors spectateur, demande le sens de l'item lexical « cache-cache » (ligne 1). Étienne profite de la situation pour impliquer Sophie, car c'est elle qui a pensé au jeu, dans la séquence d'explication en la mettant en position d'explicatrice. Contrairement aux exemples qui ont été abordés plus haut où l'apprenti-enseignant était explicataire, cette fois-ci c'est l'autre apprenant qui le devient.

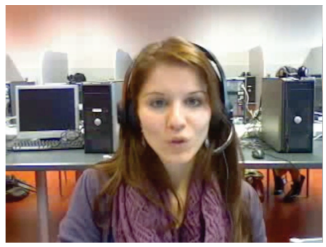
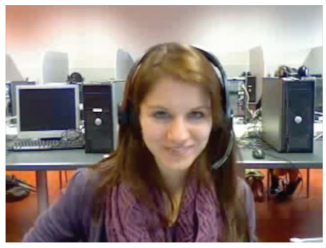
Transcription de la parole			Image du participant	Phénomènes observés
1	CNR	**qu'est-ce qui est: euh cache cache (2,2)**		Expression faciale d'incertitude (yeux plissés, l'index touche ses lèvres)
2	ETN	**EUH: [c-: explique so]phie**		
3	CNR	**[c'est un jeu nan]**		
		(1,5)		Regard pensif (yeux retournés vers le haut)[7]
4	SOH	un enfant doit (0,2) des:\ mh (2,0) mh: (2,3) se cacher/ (3,6)		
5	CNR	ah "like" euh: (1,2) ils [cour]ir		Regard pensif
6	ETN	[ouais] (0,4) en fait (1,3) **mh:**		
7	CNR	**nan/**		
8	ETN	**hm:** nan (1,1) l- (0,3) l'enfant doit: se cacher/ (0,3)		
9	SOH	oui		
10	ETN	LES enfants doivent se cacher/ (0,3) et (.) un enfant/ **(1,2) doit: trouver les autres** (1,3)		Geste iconique
11	CNR	AH oui oui oui oui (1,4) j'ai compris [maintenant\]		
12	ETN	[on ne court] pas (0,2) on ne court pas\		

Exemple 13 : Une sollicitation de l'aide d'une apprenante pour expliquer le sens

L'expression faciale de Conor témoigne de sa difficulté à comprendre le sens du mot cache-cache (ligne 1). Au lieu de l'expliquer lui-même, Étienne délègue le rôle d'explicatrice à Sophie (ligne 3). Celle-ci réfléchit avec un regard pensif et donne une courte explication (ligne 4). Conor vérifie sa propre compréhension en demandant si l'enfant doit courir (ligne 5), ce qui montre qu'il a écouté Sophie. Étienne élabore la définition donnée par Sophie et l'accompagne d'un geste iconique (lignes 8, 10 et 12). Par cette simple phrase (« explique Sophie », ligne 2), Étienne réussit à initier une séquence de négociation entre les deux apprenants. Il reste tout de même l'arbitre, car il reformule la définition donnée par Sophie, ce qui constitue aussi un moyen d'impliquer l'apprenant (Fasel Lauzon, 2014).

[7] Le floutage rend imperceptible le fait que son regard soit orienté vers le haut.

Dans les deux exemples précédents, c'était l'apprenti-enseignant qui a demandé à un apprenant de donner le mot juste ou d'expliquer le sens d'un mot inconnu. Dans l'exemple suivant, Mélissa va encore plus loin en demandant à ses apprenantes de prendre en charge elles-mêmes la totalité de la séquence d'explication lexicale. Cette séquence d'explication dont l'objet est le mot « bière » est issue de la première semaine des interactions où les apprenants parlent de la pause café et des boissons pour se détendre.

Transcription de la parole	**Image de Mélissa**	**Phénomènes observés**
1 ALJ et (.) et comment ça se dit (.) "beer"/ (.) ((rires)) **(1,0)** 2 MEL comment dit (0,5) "beer"/ ana/ (.) tu as une idée/ (0,8) 3 ANA hum euh qu'est-ce que euh j'ai pas entendu (0,6)		Regard caméra
4 MEL [alejandra/ tu peux] poser ta 5 ALJ ["beer"] (1,3) comment ça se dit "beer" i- en "french" (1,0) 6 ANA "beer"/ (1,6) 7 MEL °ouais°/ (0,6) 8 ANA bière/ (0,5) 9 ALJ [bi-] AH (.) euh 10 MEL [oui\] (0,2) c'est [ça\ (.) une] [bière\] 11 ANA [°((rires))°] 12 ALJ [bière] ((rires)) (.) **oh:**		Regard caméra et sourire

Exemple 14 : Faire collaborer ses apprenantes entre elles

Le mot « bière » fait défaut à Alejandra, qui demande une traduction du mot anglais (ligne 1). Mélissa sollicite l'aide d'Ana pour trouver la traduction française, mais cette dernière dit ne pas avoir entendu la question (ligne 3). Au lieu de prendre en charge la demande d'explication elle-même, Mélissa demande à Alejandra de répéter sa question (ligne 4). Ainsi, la séquence d'explication devient entièrement prise en charge par les deux apprenantes (lignes 5-8). Mélissa n'intervient que pour ratifier la traduction anglaise (ligne 7) et le mot juste (ligne 10). Mélissa tourne son regard vers la caméra pour impliquer ses apprenantes dans cette séquence d'explication (ligne 1) comme elle a l'habitude de le faire. Mélissa est d'ailleurs l'auteure de la moitié des séquences d'explication lexicale que nous avons repérées où l'apprenti-enseignant fait un effort manifeste pour faire collaborer les apprenants entre eux.

Pour résumer, les apprentis-enseignants peuvent impliquer les apprenants en ouvrant des espaces de travail collaboratifs dans lesquels les participants travaillent sur le lexique. Ces espaces sont souvent ouverts par un indice visuel ou verbal faisant remarquer un élément problématique dans le discours de l'apprenant. Ils peuvent être ouverts aussi par une demande d'explication, ce qui initie un travail collaboratif pour trouver la forme ou le sens d'un item lexical. Il y a *besoin* si l'apprenant a besoin de comprendre un item lexical ou d'en trouver la forme, et *évaluation* si l'apprenti-enseignant fait évaluer la forme ou le sens de l'item. Une fois trouvé, l'apprenti-enseignant peut impliquer davantage l'apprenant en lui faisant répéter la forme de l'item lexical. Il y a aussi collaboration entre les apprenants quand l'apprenti-enseignant délègue le rôle d'explicateur à l'un d'entre eux. La dimension socioaffective est maintenue par des encouragements oraux et par des sourires, surtout quand l'apprenti-enseignant partage les émotions de l'apprenant.

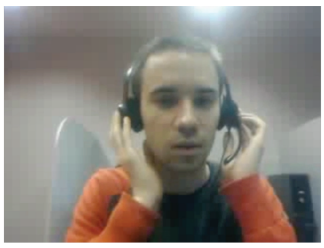
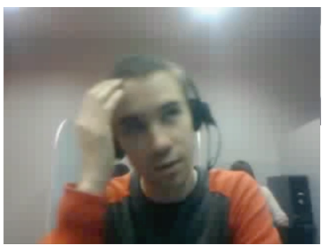
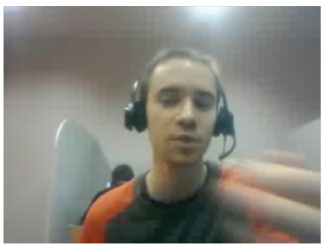
8.3 Espaces de réflexion sur la langue et sur la culture

Les sections précédentes ont montré que l'implication peut se faire en effectuant des rappels réceptifs ou productifs, ou en impliquant l'apprenant dans un travail collaboratif sur le lexique. Une forte implication cognitive peut aussi être provoquée en faisant mener des réflexions interlinguistiques et interculturelles.

8.3.1 Faire mener une réflexion interlinguistique ou interculturelle

La culture est véhiculée en grande partie par le langage, et en particulier par les mots (Galisson, 1991). Un apprenti-enseignant peut impliquer son apprenant en lui demandant de mener une comparaison interlinguistique. Par exemple, dans le chapitre 6, Adèle a demandé à ses apprenantes si la pause café existait en Irlande (exemple 16). On retrouve ce phénomène dans six séquences d'explication lexicale. On peut inciter à réfléchir non seulement sur la culture, mais aussi sur la langue elle-même.

En effet, la langue maternelle sert de point d'appui pour mieux comprendre la langue étrangère, surtout aux niveaux débutants et intermédiaires, et les comparaisons inter langagières peuvent être particulièrement enrichissantes (Giroux, 2016). Observons la façon dont Victor fait mener une réflexion inter-langagière autour de l'item lexical « vouvoiement ».

Transcription de la parole	Image du participant	Phénomènes observés
1 VIC alors: hum: (1,5) et en anglais/ (.) est-ce qu'il y a euh t- (0,2) donc (.) dans la première question/ (.) dans les deux premières phrases/ (0,3) il y avait:/ (.) euh: vous pouvez m`signer/ (.) et tu peux m`signer\ (.) donc la différence/ (0,6) c'est le: vouvoiement\ (.) tu connais vouvoiement/ (1,4) 2 LIM euh:\ qu'est-ce qu`c'est/ (1,8) (vouvoiement; normalement)/ (.) x 3 VIC <((clavardage)) Vouvoiement> (2,0) <((clavardage)) et tutoiement> (1,2) donc (.) vouvoiement/ (.) c'est:/ (0,3) le fait de dire VOUS (0,2) à quelqu'un\ (.) et pas tu\ (0,5) donc euh: (1,1) vous pouvez m`signer/ c'est du vouvoiement\ (2,6) d'accord/ (0,2) 4 LIM d'accord (.) d'accord		
5 VIC et donc (.) en anglais/ ça existe pas je crois\ (1,8) 6 LIM hm:\ **(3,4) euh: (3,8)** 7 VIC **est est-ce qu'y a une différence euh: de: vouvoiement/ et de tutoiement/ en anglais/ (0,4)** 8 LIM **euh: (0,9) je ne pense pas que: (0,8) je comprends (0,4) le mot (0,7)**		Victor touche son casque en attendant une réponse
9 VIC ah (0,6) [euh:] **vouvoiement\&** 10 LIM [hum:] 11 VIC &(0,6) alors (0,6) si par exemple (.) euh a- avec toi je hum: (0,6) je te tutoie\ (0,3) ça veut dire que je: je te dis tu (1,2) et je n- je te dis pas vous (0,7) parce que euh: (0,9) parce que c'est plus: euh: familier/ c'est plus euh:		Victor se touche la tête en en passant de l'ouverture au noyau de la séquence d'explication
(0,3) vouvoiement/ **y a une certaine distance** (.) on vouvoie/ (0,4) son: (0,7) son employeur par exemple\ (.) mais on tutoie ses amis\ (1,4)		Geste métaphorique représentant le concept de distance
12 LIM **oui (0,2)** 13 VIC **est-ce que tu comprends [mieux le mot/]**		Yeux plissés, et expression faciale d'incertitude

14 LIM 15 VIC 16 LIM	**[hum:] (1,3) euh: je pense que: (0,6) je comprends le mot (0,4) mais hum:** **d'accord\ (0,5) et donc (.) en anglais/ est-ce que ça existe/** (.) cette distinction/ (3,4) hu:m (1,5) je ne pense pas que euh: (1,1) le mot (0,4) est |u- (0,4) tilisé très souvent		Froncement de sourcils et expression faciale d'incompréhension, d'extrême difficulté

Exemple 15 : Une invitation à mener une comparaison inter linguistique

Victor demande à son apprenant s'il existe un équivalent en anglais du phénomène de vouvoiement. Il l'incite ainsi à réfléchir sur les différences entre sa langue maternelle et le français (ligne 1). Liam demande une définition du vouvoiement (ligne 2), et Victor l'explique à l'aide d'exemples après avoir écrit « vouvoiement » « et tutoiement » dans la fenêtre de clavardage. Victor relance sa demande de comparaison inter langagière (ligne 5). En attendant une réponse de Liam, il touche son casque pour mieux entendre. On pourrait dire qu'en visioconférence, le fait de toucher son casque est l'équivalent de tendre l'oreille pour mieux entendre. On ignore si Liam a le reflexe de parler plus fort quand il voit son interlocuteur toucher son casque. Liam annonce de nouveau qu'il ne comprend pas le terme (ligne 8).

Avant de commencer sa deuxième explication, Victor répète l'item lexical en question et se touche la tête. On retrouve ce comportement souvent chez Victor : quand il passe de l'ouverture au noyau d'une séquence d'explication, il se repositionne physiquement et annonce ce qu'il va expliquer. Sa deuxième explication est plus élaborée que la première, car il explique que le tutoiement est plus familier et que le vouvoiement crée plus de distance. On observe un geste métaphorique représentant le concept de distance (ligne 11). Victor accompagne sa vérification de compréhension d'une expression faciale d'incertitude (ligne 13).

A la ligne 14, Liam émet des signaux contradictoires. Il dit oralement qu'il pense comprendre, mais son visage indique clairement qu'il éprouve de grandes difficultés à comprendre le sens du terme. A la ligne 15, Victor accepte la réponse de Liam comme s'il avait compris, et relance sa demande de comparaison. La réponse de Liam à la ligne 16 témoigne de son incompréhension.

Victor implique donc son apprenant en le faisant réfléchir sur la langue, et en vérifiant la compréhension. L'effort que fait Victor pour impliquer son apprenant est louable. En revanche, on voit combien il est important de regarder les expressions faciales de l'apprenant,

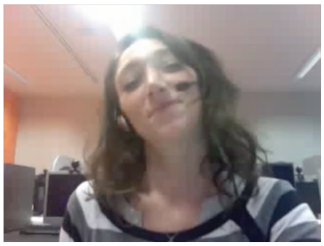
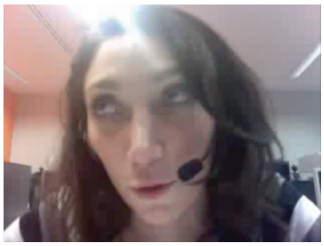
car celles-ci livrent souvent des indices d'incompréhension, indices que Victor semble ignorer. Il se peut qu'en audioconférence ou en communication par clavardage, l'apprenant aurait dit « je n'ai pas compris ». Il se peut que Liam croie que ses expressions faciales sont suffisantes pour indiquer son incompréhension. Pour les apprentis-enseignants, la webcam est donc un outil non seulement pour l'émission d'expressions faciales et de gestes, mais aussi pour la réception.

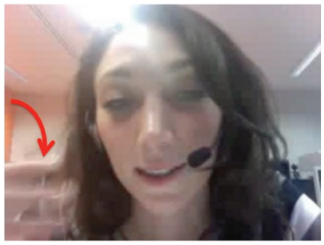
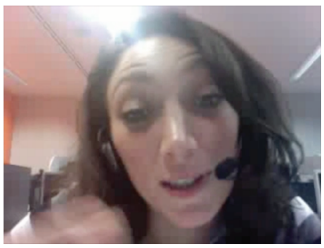
8.4 Définitions personnalisées

Plus un apprenant trouve une activité intéressante, plus il sera impliqué : l'intérêt personnel pour le sujet et pour le contenu est l'un des facteurs déterminants pour l'acquisition du vocabulaire oral (lecture publique d'un document, radio, etc.) (Nation, 2013). Dans cette ligne de pensée, nous considérons qu'un moyen d'impliquer les apprenants est de personnaliser la définition d'un item lexical en fonction des intérêts de l'apprenant, voire même de personnaliser l'item lexical. Un avantage majeur du tutorat en visioconférence réside dans le fait que l'apprenti-enseignant n'a qu'un ou deux apprenants. Il peut donc les connaître personnellement.

8.4.1 Personnaliser l'item lexical

Dans l'exemple suivant, Adèle modifie légèrement l'item lexical afin qu'il corresponde mieux aux intérêts et aux goûts personnels d'une des apprenantes.

Transcription de la parole	**Image du participant**	**Phénomènes observés**
1 ADL alors\ hum à quoi tu associes/ (.) la pause café/ (0,2) au travail\ **(1,9)** 2 ALN **quand/ (1,3)**		Éloignement de l'écran en attendant une réponse
3 ADL voulez que je répète/ (1,3) 4 ALN e[uh: s'i]l vous plaît\ 5 CTR [oui\] (0,5) 6 ADL oui\ (0,5) à quo- euh à quoi ça te fait penser:/ euh quand je te dis pause café\ (0,3) au travail\ (0,4) **tu**		Rapprochement de l'écran et regard pensif

	(0,4) qu'est-ce qui te: (.) vient en esprit\ (1,5)		Geste métaphorique, mouvements circulaires vers le bas avec la main droite
7 ALN 8 ADL 9 CTR 10 ADL	m[h:] [ou alors la **PAUSE THÉ**] pour catriona\ (0,5) ((rires)) (0,8) °hum° (0,8) catriona/ si tu as une réponse/		Haussement de sourcils et geste non-identifiable

Exemple 16 : Modification de l'item lexical

Adèle commence par poser la question qui avait été planifiée en amont de la séquence. Elle invite les apprenantes à former des associations à l'item lexical « pause café » (ligne 1). Après avoir posé la question, elle se met en retrait en attendant une réponse (lignes 1-2). Quand Allanah ne semble pas être certaine de la question, Adèle demande s'il faut qu'elle répète la question (ligne 3). Cette demande est acceptée par les apprenantes (lignes 4-5). Adèle répète la question en se rapprochant à l'écran et en faisant un regard pensif et un geste métaphorique représentant l'acte de réfléchir (ligne 6). Jusqu'ici, Adèle implique les apprenants en leur faisant créer des associations à l'item « pause café ». Puis, elle modifie légèrement l'item lexical pour qu'il devienne « la pause thé » (ligne 8).

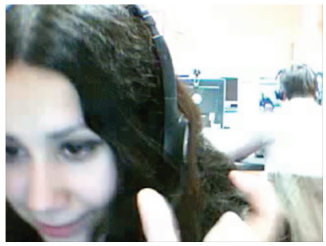
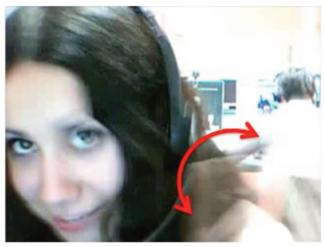
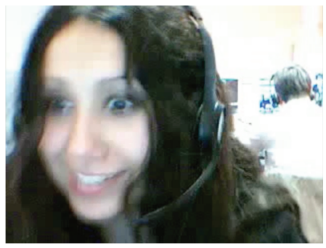
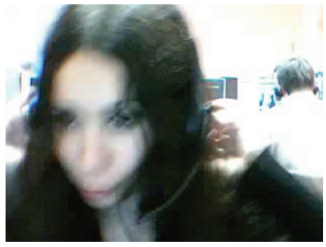
L'implication ici se joue à plusieurs niveaux. Adèle demande aux apprenantes de créer des associations à « la pause café », les invitant ainsi à participer à la construction du sens. Le sens de l'item lexical « pause café » est négocié et co-construit. Il n'est pas simplement donné par l'apprentie-enseignante. La charge d'implication est donc augmentée par la négociation du sens et par la co-construction du noyau explicatif. Adèle s'éloigne de la webcam après avoir posé sa question, en disant avec son corps : « c'est à vous ». L'*output* est ainsi encouragé non seulement par des moyens linguistiques (« qu'est-ce qui te vient à l'esprit ? »), mais aussi par des moyens visuels qui passent par la webcam. L'analyse de la conversation classique s'intéresse à la façon dont les tours de paroles sont gérés verbalement. Cet exemple montre que la webcam est utilisé pour gérer les tours de parole en visioconférence et pour encourager l'output. Une approche multimodale est donc indispensable. Puis, quand sa première demande n'est pas comprise, Adèle demande s'il faut répéter à nouveau au lieu de la donner tout de suite. Ainsi, Adèle permet aux apprenantes de s'emparer du besoin, ce qui

accroît la charge d'implication et les prépare à écouter attentivement. Enfin, afin d'impliquer Catriona, qui est souvent silencieuse, Adèle modifie l'item lexical afin qu'il s'aligne sur les intérêts personnels de Catriona qui avait mentionné qu'elle préférait le thé et non pas le café.

8.4.2 Personnaliser la définition

Dans l'exemple précédent, Adèle modifie l'item lexical afin qu'il soit adapté aux intérêts de Catriona, et pour que cette dernière puisse créer plus facilement des associations à l'item lexical. Dans l'exemple suivant, issu de la première semaine des interactions, Emilie modifie sa définition afin qu'elle contienne des éléments susceptibles de donner du sens à son apprenant et de l'intéresser. Auparavant, Emilie avait demandé aux apprenants si, à leur avis, les Irlandais étaient contents d'aller au travail. Aiden a répondu négativement en disant que les travailleurs irlandais étaient, à son avis, paresseux. Le mot « paresseux » lui a fait défaut, et il a demandé à Emilie de traduire le mot anglais « lazy ». Sept minutes et demi plus tard, une séquence d'explication lexicale autour du mot « caricature » survient. Observons la façon dont Emilie personnalise sa définition du mot « caricature ».

Transcription de la parole			**Image d'Emilie**	**Phénomènes observés**
1	EMI	alors normalement/ tu dois (.) voir une caricature/ qui s'affiche\ (0,7) est-ce que tu vois l'image\ (1,9)		
2	ADN	eh: (0,7) le mot caricature (.) eh (0,4) je ne com[prends]		
3	EMI	[oh (.)] pardon (.) alors\ (0,2) hum: une caricature/ (0,5) c'est un dessin/ (0,3) mais (.) un dessin/ où (.) TOUT est exagéré\ (0,3) tu comprends/ (0,5)		
4	ADN	non (0,7)		
5	EMI	non\ (0,9) alors\ (.) euh: (.) c'est un dessin qui est généralement (.) humoristique/ (0,2) d'accord/ (0,3) qui est censé faire ri:re/ (1,1) tu comprends/ (0,8)		
6	ADN	non <((en riant)) je suis désolé> (0,5)		
7	EMI	<((en riant)) (da-)> non n'en ne t'excuse pas c'est pas grave\ (0,6) alors (.) hum: (.) c'est- c'est la même chose que "humoristic" (.) en [anglais]		
8	ADN	[oh] "humoristique" [oh très (.) très]&		
9	EMI	["humor"\ (.) voila/ c-]		
10	ADN	&bien (.) oh oui		

11 EMI	c- c- c'est c'est un dessin eh humoristique donc qui censé faire rire/ (.) d'accord/ (0,8) et (.) pour faire rire/ (0,4) on va euh (.) bah (.) ce qu'on appelle (.) caricaturer/ c'est-à-dire/ **(0,9)** euh (0,4)		Regard pensif
	prendre **certains (.) traits de personnalité/ (0,5) prendre un** caractère/ (.) tu vois/ (0,4)		Geste métaphorique, mouvement circulaire, regard webcam.
12 ADN	[mh/]		
13 EMI	[ou u]ne personne/ (0,4) des traits de caractère par exemple/ tu disais **"la:zy" tout à l'heure/** (0,2)		Yeux grands ouverts, haussement de sourcils
14 ADN	oui/ (0,6)		
15 EMI	et bien/ (0,4) on va les caricaturer c'est-à-dire qu'on va les grossir (0,4) on va les EXAGERER\ (0,2) d'accord/ (1,0)		
16 ADN	oui/		
17 EMI	pour les rendre drôle\ (.) tu comprends/ (1,2)		
18 ADN	oui\ je comprends\ (1,0)		
19 EMI	mh par exemple hum: je te donne bah (0,4) avec l'exemple de "lazy" que tu donnais tout à l'heure/ (0,4) pour CARICATURER (0,5) quelqu'un de paresseux/ (0,7) on va le représenter/ (.) tout le temps **couché\ (.)**		Geste iconique, mains derrière la tête pour représenter une personne couchée

	avec un chapeau sur la **tête/** **(.)** et en train de boire [un verre]&		Geste iconique représentant un chapeau
20 ADN	[hm:\]		
21 EMI	&devant la mer\ (0,3) tu vois/		
22 ADN	oui oui oui (0,3) "yeah" oui (0,3)		
23 EMI	ça c'est une caricature\ (0,3) c'est comme un: (.) un stéréotype/ (0,8) tu vois/		
24 ADN	(inaud.) (.) "yeah" oui/ (.) oui je je comprends		
25 EMI	"stereotype"/ (0,4)		
26 ADN	oui [oui oui] oui oui\		
27 EMI	[voilà/] (0,6) [c'est la même] chose/(0,3) sauf que&		
28 ADN	[((rire))]		
29 EMI	&(.) les traits sont vraiment très grossis pour faire rire\ (.) tu vois/ (0,6)		
30 ADN	hm hm/		

Exemple 17 : Implication tout au long de la séquence et une définition personnalisée

Cet exemple est intéressant du point de vue de l'implication de l'apprenant pour au moins deux raisons : l'explication elle-même et les vérifications de compréhension successives. En ce qui concerne la définition, Emilie commence par donner une définition formelle (Flowerdew, 1992) (ligne 11) contenant la classe d'objets (un dessin) et une caractéristique définissante (humoristique, qui est censé faire rire). Puis, elle définit le verbe « caricaturer » (ligne 11) à l'aide d'un geste métaphorique tenant les concepts abstraits de « trait » et de « caractère » entre son index et son pouce en faisant tourner ses doigts en forme de pince et en regardant directement vers la webcam pour capter l'attention de l'apprenant. Puis, pour donner un exemple d'une caricature, Emilie reprend l'exemple de « lazy » / « paresseux » qui vient d'Aiden. Quand Emilie rappelle le mot « lazy », son visage devient animé (ligne 13). Pour rendre l'exemple vivant, Samia fait des gestes iconiques pour représenter quelqu'un qui est « tout le temps couché avec un chapeau sur la tête et en train de boire un verre devant la mer » (lignes 19-21). Ce choix d'exemple est en lien avec un mot dont Aiden a demandé la traduction auparavant. Cet exemple de caricature crée des associations et des traces mentales avec ce mot nouvellement appris. Désormais, Aiden dispose non seulement de la traduction française du mot « paresseux » mais aussi d'un exemple concret. On assiste a un épisode d'*instanciation* (Nation, 2013) où la forme linguistique du mot est relié à son sens en pensant à une expérience où on a rencontré le mot ou l'objet dans la réalité. On peut dire que c'est une instanciation simulée, car Aiden rencontre une description d'une personne paresseuse et non une vraie personne. Mais, puisque Aiden avait déjà pensé à ce mot, il est fort possible qu'il crée une image mentale d'une vraie personne paresseuse qu'il a rencontrée dans sa vie. L'exemple personnalisé choisi par Emilie

et l'illustration multimodale qui l'accompagne permettent à Aiden de créer des liens entre les mots « paresseux », « caricature », et « stéréotype ».

L'autre façon saillante d'impliquer l'apprenant dans l'explication est le nombre élevé de vérifications de compréhension que l'on observe tout au long de la séquence. Ces vérifications de compréhension permettent aux apprenants d'initier des séquences de négociation du sens si nécessaire (Pica, Young, & Doughty, 1987). Quand Emilie vérifie que l'image s'affiche sur l'écran de l'apprenant, celui-ci manifeste son incompréhension du mot « caricature » (ligne 2). Emilie s'excuse (ligne 3) avant de le définir. La première et la troisième vérification de compréhension (lignes 3 et 5) reçoivent des réponses négatives, ce qui témoigne de leur utilité dans l'identification des besoins de l'apprenant. Après chaque étape de l'explication, Emilie vérifie que son apprenant n'est pas perdu. Dix vérifications de compréhension s'observent (voir le tableau ci-contre). Elles prennent souvent la forme d'une « tag question » (Sacks, Schegloff, & Jefferson, 1974 : 718) qui servent à impliquer l'interlocuteur. Celles-ci permettent à Emilie de vérifier constamment l'état de compréhension de l'apprenant et de juger de ses besoins. Elles permettent de faire émerger, si besoin, un indicateur d'incompréhension (Varonis & Gass, 1985). Ces vérifications de compréhension sont également les garants de l'implication d'attention de l'apprenant car elles obligent l'apprenant à montrer qu'il écoute. La combinaison d'une définition personnalisée et ces multiples vérifications de compréhension témoignent de l'effort que fait Emilie pour impliquer son apprenant dans cette séquence d'explication.

N°	Ligne	Texte
1	3	tu comprends ?
2	5	d'accord ?
3	5	tu comprends ?
4	11	d'accord ?
5	11	tu vois ?
6	15	d'accord ?
7	17	tu comprends ?
8	21	tu vois ?
9	23	tu vois ?
10	29	tu vois ?

Figure 9 : Implication par « Tag questions »

Les deux exemples précédents illustrent deux façons de personnaliser l'item lexical ou la définition afin d'impliquer davantage l'apprenant. Adèle modifie un item lexical afin de faciliter la création d'associations avec celui-ci, et Emilie intègre un mot juste qui provient d'une séquence d'explication antérieure initiée par l'apprenant afin d'aider celui-ci à créer une instanciation mentale du mot nouvellement appris et afin de mieux comprendre l'explication en cours. Ce faisant, elle demande des indices de compréhension ou de non-compréhension tout au long de la séquence.

8.5 Conclusion

Dans ce chapitre, nous nous sommes efforcé d'analyser la façon dont les apprentis-enseignants impliquent leurs apprenants dans les séquences d'explication lexicale. Nous avons repéré plusieurs stratégies d'implication et les avons comptabilisées dans le tableau ci-dessous.

Stratégie d'implication	Nombre de séquences
Vérifier la compréhension d'un item lexical	91
Faire traduire le mot en langue maternelle	7
Effectuer un rappel productif	4
Faire remarquer un élément problématique	22
Faire expliquer à l'apprenti-enseignant	28
Faire évaluer une définition d'un mot juste	7
Faire évaluer une proposition d'un mot juste	55
Faire répéter la forme correcte	15
Faire expliquer à l'autre apprenant	8
Faire mener une réflexion interlinguistique	1
Faire mener une réflexion interculturelle	6
Personnaliser la définition ou l'item lexical	8

Figure 10 : Stratégies d'implication et nombre de séquences dans lesquelles elles se trouvent

La stratégie la plus présente dans le corpus est la vérification de compréhension, qui se retrouve dans presqu'une séquence d'explication sur trois. Hormis les rappels réceptifs ou productifs, nous avons démontré que les apprentis-enseignants ouvrent des espaces de collaboration dans lesquels les participants travaillent ensemble sur le lexique. Dans ces espaces, les apprentis-enseignants impliquent les apprenants en faisant remarquer un élément problématique dans l'output, en faisant expliquer, en faisant évaluer, et en faisant répéter. Un apprenant peut aussi être impliqué quand un apprenti-enseignant lui demande de réfléchir sur la langue ou sur la culture à travers le lexique. Enfin, un apprenti-enseignant peut faire un effort pour personnaliser l'item lexical ou la définition de celui-ci. Nous avons démontré que l'implication n'est pas toujours réussie : il faut par exemple laisser à l'apprenant assez de temps pour répondre à une question et ne pas donner la réponse tout de suite.

Nous avons tenté de tenir compte de la dimension socioaffective de l'implication. Nous avons constaté des encouragements et alignements socioaffectifs par des moyens verbaux, dans le cas de Samia, mais aussi des rires, sourires, et rapprochements de l'écran qui passent par la webcam. La webcam peut permettre un partage d'émotions quand les apprenants rient en même temps. Ceci fait partie du contrat didactique qui se met en place en visioconférence où il y a moins de hiérarchisation : les apprentis-enseignants se mettent parfois en position d'apprenant en demandant des mots en anglais.

A travers ces analyses, nous avons mesuré l'utilisation des canaux de Visu. Nous avons démontré notamment que la webcam joue plusieurs fonctions. Par des rapprochements de l'écran, des expressions faciales, des regards, et des gestes, les apprentis-enseignants alertent les apprenants des problèmes sur le lexique, gèrent les tours de parole, et font évaluer les mots et les définitions. Nous avons trouvé notamment que, quand il s'agit de faire évaluer une définition d'un item lexical, les apprentis-enseignant font des gestes la plupart du temps pour expliquer un mot dont ils estiment que la forme est inconnue (cf. section 8.2.3 en page 276). Il semble donc que le concept de *recipeint design* (Sacks et al., 1974) soit valable en visioconférence car les apprentis-enseignants élaborent des explications plus riches sur le plan multimodal quand ils estiment que l'item lexical est inconnu. La webcam semble jouer un grand rôle non seulement dans la focalisation sur la forme et sur le sens des items lexicaux, mais aussi dans l'implication des apprenants.

Chapitre 9 : Expliquer un même item lexical différemment

9.1 Introduction

Nos analyses précédentes nous ont permis de faire un état des stratégies déployées par les apprentis-enseignants pour se focaliser sur la forme et sur le sens des items lexicaux d'une part, et d'autre part pour impliquer les apprenants dans les séquences d'explication lexicale. Dans les chapitres précédents, aucune comparaison directe n'a été effectuée. Or, l'un des avantages de notre corpus réside dans le fait que tous les apprentis-enseignants étaient dans des conditions très similaires et suivaient les mêmes plans des séances qui avaient été élaborés collectivement. Ainsi, certains mots-clés sont abordés et/ou expliqués par tous les apprentis-enseignants dans une même séance. Cela nous permet d'effectuer alors des comparaisons interindividuelles. Peu de corpus écologiques permettent ces comparaisons.

Tellier, Michel, & Wolff (2014) ont mené une comparaison interindividuelle sur la production gestuelle de cinq enseignantes à l'école maternelle. Cette étude leur a permis de dégager des similitudes et des différences entre les enseignantes dans des situations différentes : lecture d'une histoire, explication de consignes, et gestion de la classe. Puis, les enseignantes ont été interviewées afin de déterminer si en situation d'interview la gestuelle de ces enseignantes était différente qu'en situation de classe. Les auteures ont trouvé des différences interindividuelles, ce qui corrobore le constat de Cadet & Tellier (2007) que chaque personne a un profil gestuel différent, et celui de Cosnier & Develotte (2011) qui montre que chaque personne a un profil interactif qui ne disparaît pas en ligne. A l'instar de ces chercheurs, nous proposons une étude comparative afin d'explorer les profils didactiques et multimodaux des apprentis-enseignants lorsqu'ils expliquent un même item lexical.

Les comparaisons interindividuelles que nous mènerons ici sont d'autant plus intéressantes que les apprentis-enseignants sont en formation en début de carrière, et utilisent une plateforme qui leur est relativement nouvelle. Leurs habitudes didactiques (Cicurel, 1993), sémiotiques et techniques ne sont pas encore figées et font l'objet d'un regard critique par les acteurs eux-mêmes et par les chercheurs (Guichon, 2009, 2013). Une comparaison interindividuelle sur les explications d'un même item lexical par plusieurs apprentis-enseignants permettra de dégager les différentes façons d'aborder l'item lexical, de le déconstruire et de l'expliquer.

Les avantages de ce type de comparaison viennent du fait que tous les apprentis-enseignants suivent la même formation, suivent les mêmes plans de séance, disposent des mêmes outils, et expliquent les mêmes items lexicaux à des apprenants qui sont tous de niveau intermédiaire. Pourtant, malgré tous ces éléments communs, on observe des différences dans la façon d'enseigner le vocabulaire. En effet, chaque individu est différent, chaque enseignant possède ses propres représentations de la langue et de l'enseignement, est plus ou moins à l'aise avec certaines ressources sémiotique, etc. Les apprenants, bien qu'ils aient tous des niveaux de français semblables, apprennent différemment, ce qui impacte la façon dont les apprentis-enseignants avec lesquels ils sont appariés enseignent. Puisque nos comparaisons ne portent que sur les explications d'un item lexical, les généralisations sont à faire avec prudence. Ces comparaisons ne s'appliquent donc qu'à des cas très précis.

Pour nos comparaisons, nous proposons d'analyser qualitativement plusieurs séquences d'explication portant sur le même item lexical. Nous considérons qu'une bonne explication lexicale doit être opportune, efficace, complexe, et expressive. Une séquence d'explication lexicale est **opportune** si elle arrive au bon moment, c'est-à-dire qu'elle arrive quand le besoin est ressenti par l'apprenant et quand sa compréhension est nécessaire pour la réalisation de la tâche. Elle est **efficace** si elle atteint son objectif en un temps adapté, et si la compréhension de l'apprenant est suffisante pour compléter la tâche. Elle est **complexe** si elle fournit une multitude de facettes. Pour rappel, les trois facettes principales de la connaissance d'un item lexical sont la forme, le sens, et l'usage. Puisque les apprenants n'ont que quarante minutes dans ce contexte spécifique de la visioconférence, et que chaque item lexical n'est abordé normalement qu'une fois, plus l'explication est riche, plus le temps du cours sera rentabilisé. Une explication est **expressive** si plusieurs modes sémiotiques sont utilisés lors de l'explication. Nous avons vu dans le chapitre 3 que l'usage de plusieurs modes est susceptible de renforcer la mémorisation. Enfin, une bonne séquence d'explication lexicale devrait donner lieu à une participation active de l'apprenant, à la fois pour approfondir le traitement cognitif (Craik & Lockhart, 1972 ; Laufer & Hulstijn, 2001 ; Nation & Webb, 2011), et pour respecter la vocation de la formation en ligne, à savoir la communication orale. Nous allons analyser qualitativement ces critères.

En ce qui concerne la sélection des séquences d'explication lexicale à comparer, nous avons fait le choix de partir de la quatrième semaine des interactions car toutes les séances de celle-ci ont été transcrites. Nous avons décidé de transcrire toutes les séquences de la quatrième semaine afin que le corpus puisse permettre ce type de comparaisons

interindividuelles (cf. chapitre 4). Nous disposons donc de sept interactions issues de la même semaine où tous les apprentis-enseignants suivent le même plan de séance. La sélection de cette séance nous donne plus de probabilité pour repérer un item lexical qui est traité par plusieurs apprentis-enseignants. Le tableau ci-dessous fait état de tous les items lexicaux qui ont fait l'objet d'une séquence d'explication durant la séance n°4.

Apprenti-enseignant	Adèle	Emilie	Étienne	Mélissa	Samia	Séverine	Victor
Items lexicaux expliqués	1. goûter d'anniversaire 2. je m'en fous 3. attractions 4. équilibre	1. planifier les animations 2. pot 3. tenir les délais 4. beaucoup plus 5. « not to let the others down » 6. disputes 7. « complain » 8. « takes longer »	1. paresseux 2. ultimatum 3. anniversaire à McDonald's 4. cache cache 5. Happy Meal 6. personnalité 7. autoritaire 8. clown 9. « bon très » 10. patron et clients	1. créer 2. entente 3. nominalisation 4. réduction / promotion / faire un prix 5. matière 6. inviter 7. arriver à l'heure	1. équipier 2. pense bête 3. manager 4. goûter d'anniversaire 5. aider 6. couronne	1. agressif 2. équipier 3. goûter d'anniversaire 4. hygiène 5. licencier 6. pot 7. retraite 8. systématique 9. négliger 10. s'en ficher 11. courtois 12. pense bête 13. apaiser 14. en face 15. paresseux 16. « playplace » 17. rapport 18. timide 19. boissons gazeuses 20. très souvent 21. nettoyage	1. chaises musicales 2. goûter d'anniversaire 3. jeu de rôle 4. rétroplanific-ation 5. licencier / remercier / congédier 6. employé 7. avertissement 8. entraînement 9. Père Noël 10. vêtements sportifs 11. organisation

Figure 1 : Les items lexicaux ayant fait l'objet d'une séquence d'explication durant la quatrième séance

Tous, sauf Emilie et Mélissa ont expliqué l'item lexical « goûter d'anniversaire ». Étienne a expliqué l'item « anniversaire à McDonald's ». Les apprentis-enseignants avaient comme tâche principale de mettre en place et de gérer un projet. Il s'agissait d'un anniversaire chez McDonald's que les apprenants dublinois devaient planifier en étant créatifs et en prévoyant d'éventuelles difficultés. Les apprentis-enseignants lyonnais devaient chacun jouer le rôle de manager. La consigne était : « Imagine. Tu es équipier chez McDonald's et ton manager et toi réfléchissez à une nouvelle formule pour les goûters d'anniversaire ». La compréhension de l'item lexical « goûter d'anniversaire » est donc cruciale pour le déroulement de l'activité qui s'interrompt pour résoudre les problèmes de compréhension dans les cinq cas qui seront analysés.

L'autre item lexical que nous avons sélectionné pour ce chapitre de comparaison est issu de la semaine suivante (n°5) où l'activité principale consistait à concevoir un camion-restaurant à destination d'une clientèle dite « bobo ». Quatre interactions ont été transcrites dans cette semaine, chacune contenant une séquence d'explication lexicale autour de l'item « bobo ». Nous avons déjà effectué des analyses de ces quatre séquences dans Holt & Tellier (2017) en partant d'une description des séquences d'explication par phases : ouverture, noyau, et clôture. Nous ne nous étions pas alors attachés à analyser en priorité les facettes de la connaissance de l'item lexical faisant l'objet d'une focalisation. Dans ce chapitre, nous approfondirons les analyses déjà effectuées et étudierons ces mêmes séquences sous un nouvel angle.

Les deux items lexicaux cités ci-dessus portent une charge culturelle partagée importante (Galisson, 1991). « Goûter » n'a pas la même signification en France qu'en Irlande. Pour les Anglo-Saxons, un goûter (*a snack*) est l'acte de grignoter entre les repas, que ce soit le matin, l'après-midi, le soir, à la maison, dans la rue ou dans les transports. Pour les Français, le goûter est plus ritualisé, se faisant généralement après l'école entre 16h et 17h et étant surnommé « le quatre heure ». Le mot « bobo » est un mot-valise crée par la troncation des mots « bourgeois » et « bohème » et n'a pas d'équivalent exact en anglais. Ce concept est même très français pour les originaires d'autres pays francophones, comme en témoigne l'échange ci-dessus qui est issu du débriefing de la cinquième séance. Victor, qui est québécois, commente le caractère « très français » de l'item lexical :

Apprenti-enseignant	Parole
Séverine	Et on dit bobos au Canada, au Québec ou comment ça se passe ?
Victor	Oui mais c'est un concept très français pour nous quand même.
Séverine	Très français.
Victor	Ça existe mais c'est pas aussi…

Transcription 1 : Discussion des bobos lors du débriefing (Séverine et Victor)

L'item lexical est variable non seulement selon les cultures, mais selon les individus aussi. Emilie, dont la séance n'a pas été transcrite, en dit autant lors du débriefing :

Apprenti-enseignante	Parole
Emilie	Mais c'est précis finalement comme (terminaison) parce que chaque fois j'entends ce mot employé de partout mais j'arrive jamais à mettre une définition précise dessus. J'ai l'impression que ça varie selon toutes les personnes.

Transcription : 2 Discussion des bobos lors du débriefing (Emilie)

La forte charge culturelle de cet item lexical rend intéressante une comparaison des différentes façons de l'expliquer.

Nous commencerons par analyser les explications de l'item « bobo ». Puis, nous comparerons les explications de l'item « goûter d'anniversaire.

9.2 Bobo

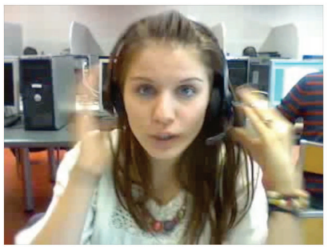
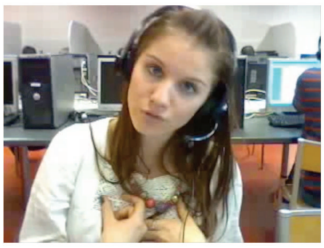
Dans cette section, nous analyserons par critère les explications de l'item lexical « bobo ». Nous commencerons par analyser l'opportunité des séquences, en regardant les phases d'ouverture ainsi que les lignes précédentes. Puis, nous analyserons les critères *complexité* et *expressivité* ensemble. Enfin, nous discuterons de l'efficacité des explications en considérant le temps nécessaire pour chaque explication. Pour cette tâche, les apprenants devaient jouer le rôle d'un expert commercial qui devait étudier le marché avant de concevoir un camion restaurant à la Croix Rousse à Lyon, un quartier dans lequel logent de nombreux bobos. Des explications lexicales sur le mot « bobo » ont eu lieu lors de cette activité. Lors de la préparation de la séance, la meilleure façon d'expliquer l'item lexical avait été discutée. Des traces de planification sont observables dans certaines explications.

9.2.1 Opportunité

Dans cette section, nous analysons la façon dont la séquence d'explication lexicale est abordée afin de voir si elle arrive au bon moment. Nous considérons qu'une séquence d'explication lexicale arrive au bon moment si le besoin de connaître l'item faisant l'objet de l'explication est ressenti par les apprenants (Laufer & Hulstijn, 2001), et s'il y a focalisation sur l'item, car on apprend ce sur quoi on se focalise (Barcroft, 2006). Afin d'analyser ces critères, nous allons regarder les quelques secondes d'interaction qui arrivent avant le début de l'explication elle-même. Ces moments nous renseignent sur le passage entre la focalisation sur les consignes de l'activité et la focalisation sur la langue.

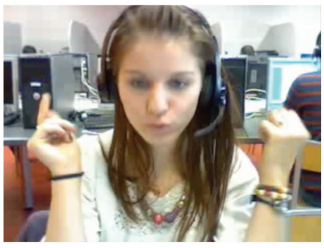
9.2.1.1 Transitions naturelles et lisses

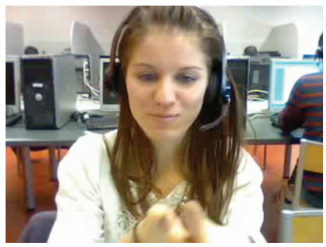
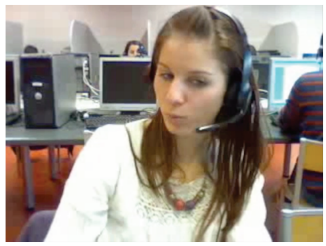
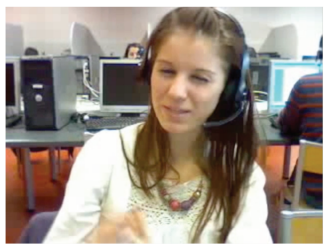
La transition vers la séquence d'explication lexicale doit sembler naturelle et ne pas être trop scolaire et forcée. Si le *focus on form* est implémenté correctement, c'est-à-dire sans nuire à la communication et à la tâche, l'acquisition est favorisée et facilitée (Long, 1991). Mélissa et Adèle parviennent à faire glisser la séquence d'explication lexicale dans l'activité de manière naturelle. Après avoir expliqué les consignes, Mélissa laisse la parole à ses apprenantes :

Transcription de la parole	Image du participant	Phénomènes observés
1 MEL donc/ ça va être d'abord l'étude du marché:/ donc les clients:/ les horaires d'ouverture:/ (0,6) et cetera\ (1,5) donc je vous écoute/ **j'écoute vos questions/** et puis moi je vais jouer/& 2 ANA <((clavardage)) quest ce que vous voulez vendre et a qui? les jeunes les ouvriers lesenfants?>		Regard caméra
3 MEL &(0,2) le rôle/ (0,3) **d'un professionnel/** (0,9) d'une cuisinière\ (1,7) alors\ (.) ana est très rapide/ ((rire)) (0,3) [((rire))] alors\ 4 ALJ [((rire))]		Regard caméra et geste déictique vers soi

Transcription 3 : Démarrage de l'activité

On remarque que, quand Mélissa explique les consignes de l'activité, elle regarde la caméra et utilise son corps. On retrouve souvent ce comportement chez Mélissa. En effet, quand elle sollicite une réponse des apprenantes, elle regarde directement vers la caméra. Immédiatement après avoir reçu les consignes, Ana, par le biais du clavardage, demande le type de clientèle souhaité pour le camion restaurant. Ainsi se présente une belle opportunité pour introduire l'item lexical « bobo » :

Transcription de la parole	Image du participant	Phénomènes observés
1 MEL qu'est-ce que je veux vendre/ et à qui/ (1,2) aux jeunes/ aux ouvriers/ aux enfants/: alors\ **je vise/** (1,1)		Geste métaphorique, regard vers le haut
plutôt/: une clientèle/ (1,8) euh: (0,3)		Geste déictique vers le haut

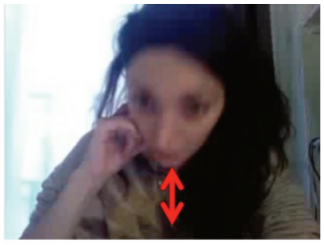
	j`sais pas si **ana m'entend/** (0,4)		Geste déictique vers l'écran
	je vise une clientèle euh (.) un peu/ (.) bobo\ **(0,8)**		Éloignement de l'écran, tête inclinée
2 ALJ	**j`sais pas si vous connaissez** ce terme/ (0,6) bo[bo/] [non]		Geste métaphorique, yeux plissés

Transcription 4 : Une difficulté de compréhension anticipée

Par des moyens différents, Mélissa fait en sorte que le besoin de l'explication se fasse ressentir et met en scène l'introduction de l'item lexical en créant du suspens. Mélissa commence par répéter le verbatim de son apprenante afin de capter l'attention de celle-ci et de donner une raison d'être à l'explication qui va suivre. A l'aide d'un geste métaphorique et d'un déictique, Mélissa montre le fait qu'elle vise une certaine clientèle. Puis, elle vérifie le fonctionnement du canal audio. Elle reprend le mot « clientèle » et entoure la première prononciation du mot « bobo » de pauses. Ainsi, il est clair que Mélissa est en train d'introduire un certain type de clientèle, et on sait pourquoi, et qu'il s'agit d'une information importante car elle vérifie le fonctionnement du canal audio. Il se peut qu'elle vérifie aussi que son apprenante est bien à l'écoute et attentive. Après avoir introduit l'item lexical, Mélissa vérifie la compréhension auprès des apprenantes. Cette introduction arrive au bon moment car elle répond à un besoin de l'apprenante (« Qu'est-ce que vous voulez vendre et à qui ? »). La difficulté semble être anticipée par Mélissa, comme en témoignent les yeux froncés qui indiquent une difficulté potentielle de compréhension, ainsi que son énoncé « je ne suis pas sûre que vous connaissez ».

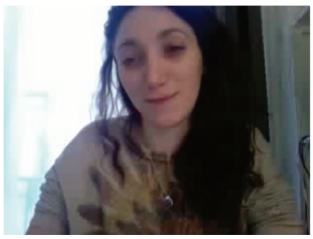
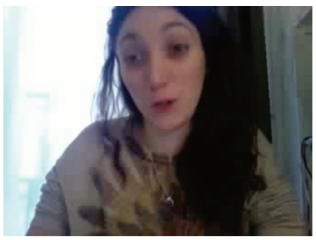
Adèle a la même chance que Mélissa de pouvoir introduire un nouvel item lexical à un moment opportun. Ici, l'introduction du mot « bobo » s'enchaine sur une séquence

d'explication lexicale autour de l'item « cible marché », qui est recherché et enfin trouvé par Alannah. Le début de cette séquence a été transcrit dans l'exemple 9 du chapitre précédent.

Transcription de la parole	Image du participant	Phénomènes observés
1 ALN <((clavardage)) cible marche?> (1,2) 2 ADL cible marché\ (0,4) ah\ (.) donc la cible marché ça va être (.) la clientèle/ (1,2) 3 ALN "yeah target market" (0,3) **"so"** (1,2) "so" (0,3) [on peut] euh/ 4 ADL j[e veux/] (0,2) je v- (2,5)		Hochement de tête

Transcription 5 : Un item lexical propice à une transition vers les bobos

Adèle saisit l'opportunité pour parler du type de clients auxquels est destiné le camion restaurant :

Transcription de la parole	Image du participant	Phénomènes observés
1 ADL je veux une clientèle euh: (0,3) bobo/ (0,5) tu connais bobo/ (1,1) <((clavardage)) bobo> **(4,5)**		Tête inclinée en attendant une réponse, sourire progressif
tu connais/ (0,5) 2 ALN non (1,2)		Haussement de sourcils

Transcription 6 : Une transition lisse et une vérification de compréhension

Après avoir introduit le mot-clé, Adèle vérifie la compréhension et attend longuement une réponse avec un sourire et un haussement de sourcils. Le besoin de comprendre le nouvel item lexical se fait donc ressentir. On assiste à une nouvelle introduction opportune qui s'intègre convenablement dans la discussion sur la clientèle et qui a été abordée par l'apprenante.

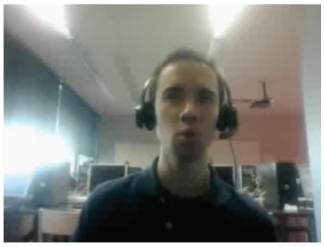
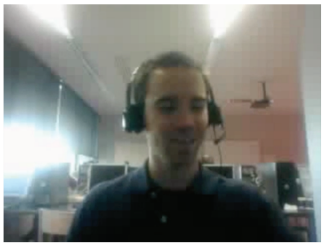
9.2.1.2 Une introduction orientée

Victor, contrairement à Mélissa et à Adèle, n'a pas la même opportunité d'introduire l'item lexical en répondant directement à une question posée par son apprenant. Liam, au lieu de demander le type de clients recherché, demande le type de nourriture qui sera vendu.

Transcription de la parole		
1	LIM	euh: (0,4) quelles sortes euh: (0,4) alimentaires (0,5) a- avez-vous/ (1,8)
2	VIC	alors j'ai prévu faire hum: (0,3) des euh: burgers:/ (0,7) mais de très bonne qualité avec de la viande: bio: et des aliments locaux:/ (0,3) donc euh:

Transcription 7 : La nourriture que mangent les bobos

En effet, Victor effectue sa transition vers les bobos en insistant sur la qualité de la nourriture qui devra être vendue. Puis, Victor demande à Liam si le mot « bobo » est connu :

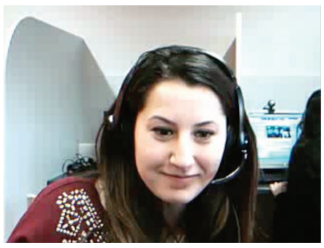
Transcription de la parole			Image du participant	Phénomènes observés
1	VIC	comme c'est^une clientèle hu:m qu'on appelle bo**bo/ (.)** est-ce que tu connais euh: (.) le mot bobo/		Regard pensif
2	LIM	non (1,0)		
3	VIC	**haha**		Victor pose sa bouteille, sourit, et se repositionne sur sa chaise.
4	LIM	**((rires))** **(1,0)**		

Transcription 8 : Une transition moins lisse et une vérification de compréhension

Contrairement aux deux exemples précédents, la transition vers le mot-clé est guidée par l'apprenti-enseignant plutôt que par l'apprenant. Victor aborde le mot-clé à ce moment car le type de clients a un effet sur le type de nourriture vendu. Ceci rend la transition naturelle et logique. Une définition du mot « bobo » pourrait être inférée par la transition de Victor : un bobo est une personne qui aime les burgers de bonne qualité.

9.2.1.3 Une transition moins fluide

Samia et ses apprenants discutent du camion-restaurant avant d'aborder le type de clientèle. Ils pensent à plusieurs idées, qui vont de la publicité sur des réseaux sociaux à des offres spéciales, comme par exemple deux hamburgers pour le prix d'un. Ils discutent de tout cela sans connaître le type de clientèle. La discussion sur le mot-clé arrive brutalement après coup :

Transcription de la parole			Image du participant	Phénomènes observés
1	SAM	donc voilà\ (0,2) d'accord\ (2,3) ah j'ai oublié d`vous envoyer les: l- les: (0,2) le TYPE de clients en fait\ (.) j`vous envoie la photo de: (0,4) des cliENTS (.) euh qui habitent dans c`quartier\ (0,8) qui vont venir manger\ **(4,8)**		Regard vers l'écran et demi sourire
2	ANG	ah:\ d'accord\ (3,2)		
3	SAM	on les appelle les bobos\ (2,1) ((rire)) **(0,7)**		Sourire
4	ANG	**pourquoi: est-ce que:** (0,5) on les appelle les bobos\ (0,9)		

Transcription 9 : Une introduction brutale du mot « bobo »

La réponse d'Angela (ligne 2) n'arrive que cinq secondes après l'introduction du mot « bobo ». L'apprenante semble donc avoir du mal à percevoir l'intérêt de ce mot. Sans le définir, Samia dit « on les appelle les bobos » (ligne 3) et attend un peu plus de trois secondes pour que l'apprenante lui demande une définition (ligne 4). Cette séquence d'explication est donc moins opportune que les trois autres car elle arrive après que la conception du camion restaurant est discutée. La réponse tardive et surprise d'Angela témoigne du caractère inopportun de ce début de séquence.

Nous avons vu deux débuts de séquences d'explication lexicale très opportunes, qui s'insèrent facilement et naturellement dans la discussion du camion restaurant. Mélissa et Adèle ont la chance de pouvoir aborder le mot-clé « bobo » comme réponse naturelle et attendue à des questions posées par leurs apprenantes. Bien que l'apprenant de Victor (Liam) ne pose aucune question directement concernant la clientèle, celui-ci arrive à réorienter naturellement la discussion vers les clients car le type de clients influence la nourriture qui devra être vendue. Samia semble avoir oublié d'aborder le mot-clé « bobo » et essaie de l'aborder après coup, ce qui est maladroit. Tous sauf Samia vérifient la compréhension explicitement.

9.2.2 Complexité et expressivité

Une explication lexicale est complexe si plusieurs facettes de la connaissance d'un item lexical sont abordées. Cela veut dire qu'il y a focalisation sur la forme, sur le sens, et sur

l'usage de l'item lexical. Il existe des arguments pour et contre la mise en œuvre d'explications complexes. Ellis (1995) a étudié l'apprentissage de mots par négociation, et a trouvé que plus une définition est riche, complexe, détaillée et prolixe, moins elle est efficace pour la mémorisation. Chaudron (1982) a trouvé que les définitions longues et compliquées induisent l'apprenant en erreur. Bogaards (1994) est du même avis : les définitions doivent être simples et utiliser un vocabulaire restreint. Ce n'est pas grave si l'apprenant n'acquiert qu'une connaissance partielle d'un item lexical en classe de langue, car la connaissance sera complétée au fur et à mesure lors de rencontres ultérieures (Nation, 2013).

En même temps, une explication complexe de l'item lexical « bobo » se justifie à la fois par la nature de l'activité et par l'item lui-même. L'activité nécessite une compréhension profonde du sens du terme « bobo » car l'étude de marché implique que l'on obtienne une connaissance intime des clients et de leurs goûts. Si une explication sur les composants d'un mot est facultative la plupart du temps, les composants des mots-valises comme « bobo » font à plus forte raison partie de la connaissance de l'item lexical et ne peuvent donc être ignorés. Concernant la pratique de la langue, les apprenants vont séjourner en France, et les connaissances culturelles et langagières transmises par les apprentis-enseignants ont une valeur concrète. Une explication complexe s'impose et celle-ci doit aborder toutes les facettes de l'item lexical : sa forme, son sens, et son usage.

Une explication lexicale est expressive, selon nous, si l'apprenti-enseignant tire partie des affordances de l'environnement pour enrichir l'explication. Comme nous l'avons évoqué dans le chapitre 3, la distribution des informations sur plusieurs modalités a au moins deux avantages : la mémorabilité est augmentée et le temps de l'explication est diminué (Kelly, McDevitt, & Esch, 2009).

9.2.2.1 Une transition expressive

Ce qui est très intéressant dans ce cas c'est que Victor, à la ligne 3 (cf. *supra* en page 305), marque clairement la transition entre les deux premières phases de la séquence d'explication. Quand il passe de l'ouverture au noyau de la séquence, il sourit, rit, pose sa bouteille et se repositionne sur sa chaise. Van Leeuwen (2005) s'appuie pour Scheflen (1964) pour dire que les frontières entre les phases multimodales sont souvent marquées visuellement, par exemple par un changement de posture. La webcam montre donc bien les changements de phases.

9.2.2.2 Une explication lexicographique peu expressive

A la suite de cette transition, Victor n'utilise plus beaucoup la webcam. Pour cette raison, nos transcriptions ne prennent pas en compte ce canal :

		Transcription de la parole
5	VIC	c'est (0,3) hum: (2,9) <((clavardage)) Un bobo.> (0,5) et en fait/ (.) c'est^une contraction de deux mots (0,7) donc c'est euh: (0,3) bobo pour (0,5) bourgeois/ (0,8) bohème\ <((clavardage)) bourgeois-bohème> (1,8) et (0,4) en fait/ (0,3) c'est^une hum: c'est un peu péjoratif/ (0,8) c'est^une euh classe sociale hum: souvent ce sont des cadres (0,6) avec euh:\ (0,4) qui sont euh: (.) un peu à gauche/ (0,4) mais qui ont euh: (0,4) un bon revenu (0,9) donc euh ce sont des gens qui ont hu:m (.) un pouvoir d'achat important/ (0,5) mais qui ont une con- une conscience euh sociale/ environnementale/ et cetera/ et cetera\ (2,2)
6	LIM	d'accord/
7	VIC	est-c`que vous avez un équivalent euh: (.) en irlande/ (0,5)
8	LIM	hum: (1,7) je ne sais pas (0,3) hum: (1,2) je ne suis pas sûr (0,2) hum (2,8) euh:
9	VIC	s- si je t'envoie une photo est c`peut-être que ça peut t'aider (2,8)
10	LIM	hum: (3,3)
11	VIC	est-ce que (.) t'arrives à bien voir/ (0,2)

Transcription 10 : Une définition lexicographique du mot « bobo »

Victor aborde toutes les facettes de la connaissance du terme : la forme, le sens et l'usage. Il commence par se focaliser sur la forme, en disant que c'est une « contraction de deux mots » (ligne 5). Il donne par clavardage à la fois le mot-valise en entier et ses composants. Ceci laisse à penser que Victor estime que ces deux mots devraient être connus par l'apprenant. Puis, il donne une définition lexicographique en décrivant la classe sociale à laquelle appartiennent les bobos ainsi que leurs orientations politiques (ligne 5). Il insère une mention sur l'usage de l'item lexical, en disant que « c'est un peu péjoratif » (ligne 5). A la ligne 7, Victor tente à la fois d'impliquer l'apprenant dans la séquence et de vérifier sa compréhension. La réponse de Liam à la ligne 8 montre que la définition de Victor n'a pas été efficace. Victor envoie alors une photo pour se focaliser davantage sur le sens de l'item lexical. Ni la focalisation sur les composants du mot, ni la définition lexicographique, ni la photo ne parviennent à faire comprendre le sens de l'item lexical à Liam, car celui-ci dit qu'il ne comprend toujours pas :

Transcription de la parole	
12 LIM	euh: (0,5) j- (.) je le vois mais euh (1,0) qu- qu'est-ce que c'est euh un (0,3) bobo hum
	(4,4)
13 VIC	c'est hu:m:\ (1,7) c'est^une personne qui gagne bien sa vie/ (0,3) donc qui fait quand même euh: qui:\ (.) qui a un bon revenu/
	(0,5)
14 LIM	[ah:] (0,3) [oui]
15 VIC	[ma]is qu[i]/ hu:m\ (0,5) mais qui trouve important de d'ach`ter des aliments euh bio:/ hu:m (,6) qui est hum: sensibilisée à l'environnement:/ euh qui vote plutôt à gauche quoi
	(0,5)
16 LIM	oui euh:
	(3,7) ils ont des euh beaucoup de choses (0,5) euh beaux (1,0) [et] et jolies&
17 VIC	[oui]
18 LIM	&(1,1)
	oui
19 VIC	ils (.) oui oui c'est ça [ils]&
20 LIM	[oui]
21 VIC	&aiment bien être euh de ils aiment bien: être bien habillés en fait
22 LIM	d'accord

Transcription 11 : Une nouvelle définition lexicographique du mot « bobo »

Victor relance sa définition lexicographique (lignes 13-15) en se focalisant sur le sens. Victor semble préférer se focaliser plus sur le sens de l'item lexical que sur la forme. La focalisation sur la forme sert d'introduction, tandis que le reste de l'explication est une définition lexicographique, manifestement trop complexe pour être comprise par son apprenant. L'usage n'est traité que très brièvement, quand Victor dit que le mot est péjoratif, mais on peut se demander si cette facette de la connaissance de l'item lexical est vraiment importante si le sens du mot n'est pas compris. Victor aborde tout de même les trois facettes de la connaissance de l'item lexical : la forme, le sens, et l'usage.

Victor n'utilise pas la webcam sauf pour la transition des phases. Pour cette raison, l'explication n'est pas très expressive. Victor maîtrise tout de même l'usage du clavardage. Quand il écrit « Un bobo. » (ligne 5), ceci remplace sa voix et devient complémentaire. Son langage est distribué entre ces deux canaux de Visu sans aucune redondance. Puis, quand il écrit « bourgeois-bohème » (ligne 5), le langage écrit est redondant au langage oral et sert de renforcement multimodal. L'apprenant a donc accès aux formes orale et écrite en même temps. Quand Victor se focalise sur le sens et sur l'usage, il utilise exclusivement le canal audio, ce qui limite l'expressivité de l'explication.

9.2.2.3 Une focalisation principalement sur la forme

Après avoir envoyé une photo, Samia se focalise uniquement sur la forme de l'item lexical :

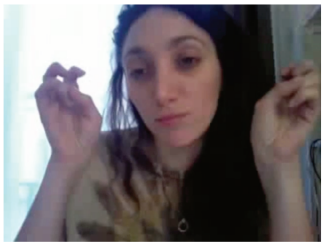
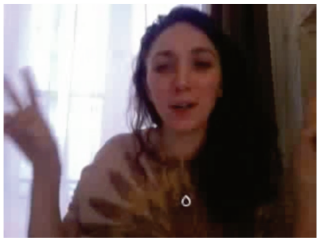
Transcription de la parole		
3	SAM	on les appelle les bobos\ (2,1) ((rire)) (0,7)
4	ANG	pourquoi: est-ce que: (0,5) on les appelle les bobos\ (0,9)
5	SAM	euh:: je sais (0,2) pas:/ pourquoi/ exactement/ (0,3) parce qu'en fait c'est des bourgeois/ (1,5) (parce)qu'on app[elle des bourgeo]is/
6	ANG	[°ah: d'accord°]
		(1,6)
7	SEN	°bobo°
		(1,1)
8	SAM	°bourgeois°/ (1,2) bohème\ (1,6) <((clavardage)) bourgeois-bohèmes> (1,1) donc on prend les: (.) les deux premières lettres/ (0,7) les: les deux premières lettres de chaque mot/ et ça donne/ (.) bobo\ <((clavardage)) bobo> (4,7)
9	ANG	°d'[accord°]
10	SEN	[bourgeois °b]ohême°
11	SAM	bourgeois bohême/ (.) et on prend les deux premières lettres/ (0,8) et ça donne/ (.) bo/ (0,4) bo\ (0,2)
12	SEN	d'accord [xx]
13	SAM	[((rire))] (0,7) [((rire)) d'accord/]
14	ANG	[j'ai mis] bobo/ (0,3) sur google et le première chose que je vois est: qu'il y a un restaurant à dublin/ (0,6) qui s'appelle bobo/ (0,4) et: (.) il vend des hamburgers\ (0,5)
15	SAM	((rire)) (0,6) voilà\ (1,0) ((rire)) (1,1) ((rire)) donc hum: (1,1) maintenant qu'on a discuté d`tout ça\

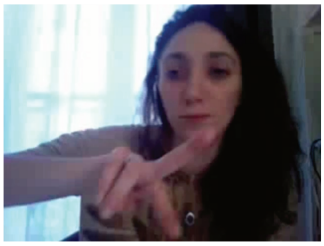
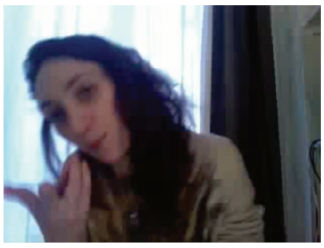
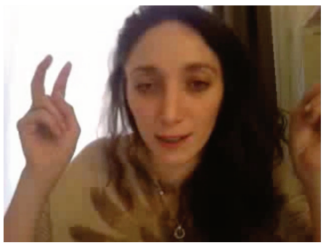
Transcription 12 : Une explication focalisée sur la forme du mot « bobo »

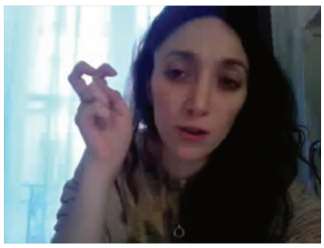
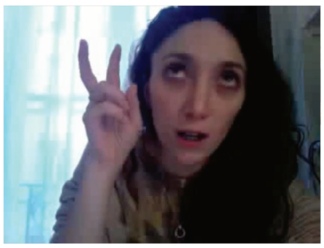
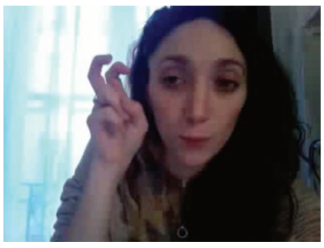
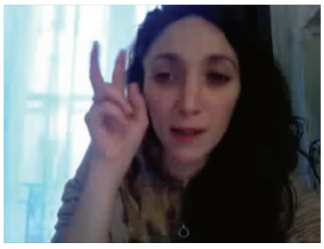
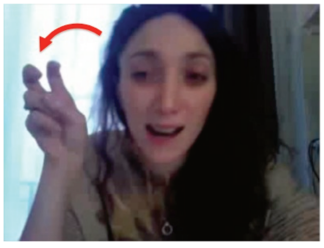
Samia est surprise par la demande d'explication de son apprenante (ligne 4). Elle décrit alors les composants du mot-valise (lignes 5-8) à l'oral et par écrit. Comme on a vu chez Victor, le clavardage est redondant à sa voix. Samia explique la construction du mot-valise (ligne 8), contrairement à Victor qui liste les composants sans expliquer la façon dont ils se combinent. Elle ne donne aucune explication, ni du sens, ni de l'usage du terme. En l'absence d'une explication du terme, Angela cherche sur Internet le sens du terme (ligne 14), et l'exemple qu'elle trouve est validé par Samia (ligne 15). Le niveau d'expressivité est bas, car Samia n'utilise pas la webcam, et n'utilise le clavardage que pour répéter ce qu'elle dit oralement.

9.2.2.4 Une focalisation multimodale la forme et sur l'usage

Adèle se focalise sur la forme, sur le sens et sur l'usage. Contrairement à Victor et à Samia, Adèle utilise non seulement sa voix et le clavardage, mais aussi la webcam et son corps pour se focaliser sur la forme de l'item lexical :

Transcription de la parole	**Image du participant**	**Phénomènes observés**
3 ADL hum\ (0,3) bobo/ c'est/ (.) hu:m (1,4) bourgeois (0,7) bohème/ (0,4) <((clavardage)) bourgeaois bohême> (1,5) en france/ (.) [c'est pareil que] 4 ALN [oh (.) okay okay\] bourgeois bohème okay\ (0,3)		
5 ADL <((clavardage)) hipsters> (0,7) "hipsters"\ (1,2) mais (.) en france/ (.) on dit/ [hu:m\] 6 ALN [c'est p]as [des b c b g\] 7 ADL **[pour dire "hipster"/] on dit/ bobo** (3,5) nan\ (.) <((en riant)) xx[x]> 8 ALN [mh] (0,2)		Guillemets gestuels
9 ADL b c b [g:/ c'e]st:\& 10 ALN [okay] 11 ADL &(0,3) c'est [plus]& 12 ALN [((rire))] 13 ADL &eu[h (0,2) **"work]ing"& (0,2)** 14 ALN **[(inaud.)]** 15 ADL **&"working [person"]** 16 ALN ["no no"] euh je [je s- euh] je sais& 17 ADL [mais: ((rire))]		Regard pensif, Adèle touche les épaules.
18 ALN &[je sais\] 19 ADL ["you know"] "i'm a **working man:**" (0,3)		Guillemets gestuels
20 ALN non mais c'est un (0,2) "it's a bit" 21 ADL tu sais le bobo/ (0,2) 22 ALN "yeah yeah yeah" (0,2) non je comprends (0,3) mais c'est (0,5) hum: 23 ADL mh (0,8) tu com- (0,8) 24 ALN euh: je sais/ bourgeois bohème/ mais pas des bobos\ (3,4) 25 ADL donc en fait/ (.) bobo/ (0,4) **tu vois le mot/ euh bourgeois/** (0,5) [et bohème/] 26 ALN c'[est un nouveau mot] pour moi\ 27 ADL ça commence les deux par b o/ (0,4) et (1,0)		

28 ALN 29 ADL 30 ALN	oui oui\ tu vois regarde (0,6) bour[geois/ tu as b] b b o/& ["yeah" (.) bo\]		
31 ADL	&(0,8) et b o (.) bohème\ (0,2) do[nc **bo/] (0,4)**		Battement tenu lors de l'énonciation de la première syllabe
32 ALN	**bo** [non non non e]uh\ (0,5) eh b o b (oui)\ (0,2) c'est ça (1,1)		Battement tenu lors de l'énonciation de la seconde syllabe
33 ADL 34 ALN 35 ADL	ça vous servira beaucoup en france/ pa`ce qu[e: les gens utilisent]& [okay le x marché/] &**beaucoup** (1,0)		Haussement de sourcils et geste vers le haut
	ils disent beaucoup ah **c'est** **un bobo euh** (2,2) ça ça ça		Geste emblématique et regard hors champ
	veut souvent dire/ **[des gens qui: qui ont]**		Guillemets gestuels
36 ALN 37 ADL	[euh: don- euh]: (2,5) (0,6) donc euh (0,7) premièrement/ on (.) on peut euh: (0,8) on peut commencer/ euh: (.) à travailler/ au: (1,0) à vendre/ (0,9) eh: (1,2) nos produits/ dans^euh: (0,5) les quartiers/ (0,3) eh (0,8) où euh les bobo habitent/ (1,1) oui		
	((67 secondes de discussion sur la mise en place du camion restaurant y compris une séquence d'explication lexicale sur l'item « centre financier »))		

38 ADL	oui/ (1,2) et ils veulent manger (.) bien/ (0,3) maintenant\ (1,5) hum (0,2) 39 ALN et [vite (.) aussi\] 40 ADL [bobo c'est] ça\ c'est: (1,1) c'est avoir de l'arg**ent/**		Guillemet gestuel
	(0,8) euh:		Guillemet ouvert
	travailler beaucoup/ (0,4)		Guillemet fermé
	mais: (.) aimer		Guillemet ouvert
	l'art:		Guillemet éloigné de la tête et fermé
	et la vie **cool quand même** (1,6) 41 ALN o:kay\ (.) "yeah"\		Geste iconique voulant dire « cool »

Transcription 13 : Une focalisation multimodale sur la forme

Adèle, comme Victor et Samia, commence par décomposer le mot-valise à l'oral et par écrit (ligne 3). Elle le décompose ultérieurement dans la séquence à l'oral et en utilisant

son corps (lignes 29-31). Contrairement à Samia, qui ne se focalise pas du tout sur le sens, et Victor, qui emploie une définition lexicographique complexe, Adèle se focalise sur le sens à l'aide d'une traduction (ligne 5) et d'une définition plus simple (ligne 40). Adèle se focalise également sur l'usage de l'item lexical en abordant son utilité et sa fréquence (lignes 33-35).

Adèle se montre très expressive durant cette séquence d'explication. En plus d'utiliser à la fois sa voix et le clavardage pour donner une traduction, quand Adèle prononce le mot « hipster », elle utilise des guillemets gestuels. Ce geste méta-explicatif complète la parole en transmettant des informations qui sont absentes du canal audio. On sait grâce à ce geste que c'est une traduction et non pas une définition française. Puis, quand Alannah demande si « bobo » égale « BCBG » (ligne 6), Adèle emploie à nouveau des guillemets gestuels pour indiquer que sa parole provient d'un discours rapporté (ligne 19)[1], et quand elle tente une nouvelle explication qu'elle abandonne aussitôt (ligne 35). Plus d'une minute plus tard dans la séquence, Adèle utilise à nouveau des guillemets gestuels pour résumer sa définition des bobos, en utilisant ses gestes pour rythmer la parole (ligne 40). Adèle exprime le sens du mot « cool » à l'aide d'un geste et d'une expression faciale (ligne 40).

Adèle est surtout expressive quand elle se focalise sur la forme de l'item lexical. Quand elle décompose le mot-valise au début de la séquence, elle utilise sa voix et le clavardage, comme les autres apprentis-enseignants. Mais quand Alannah indique à la ligne 24 qu'elle ne comprend toujours pas le sens de l'item lexical, Adèle réexplique la composition du mot-valise en utilisant cette fois-ci son corps. A la suite de cette explication orale de la composition du mot-valise (ligne 29), Adèle associe un geste et un mouvement du corps à chaque syllabe, *bo* et *bo* (ligne 31). On trouve ici un bel exemple d'un geste d'information phonologique (Tellier, 2008a) qui est rendu très visible grâce à de grands mouvements du corps, tout en restant dans le champ de la webcam.

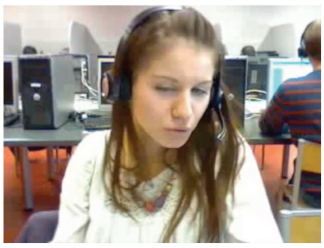
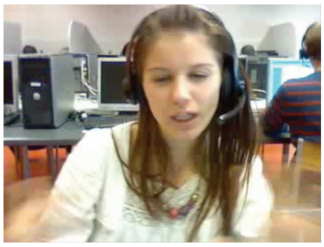
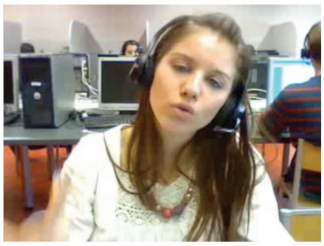
On observe également une grande expressivité quand Adèle se focalise sur l'usage de l'item lexical. Quand elle parle de l'utilité et la fréquence du terme (ligne 33-35), Adèle fait un geste métaphorique qui correspond temporellement et sémantiquement au mot « beaucoup » (ligne 35). Par ce geste et haussement de sourcils, elle renforce le mot « beaucoup » afin d'insister sur l'utilité et l'importance de l'item lexical qui fait l'objet de la séquence d'explication. L'insistance sur l'utilité de l'item lexical est susceptible d'encourager sa mémorisation par les apprenantes. Enfin, Adèle anime par un geste emblématique le

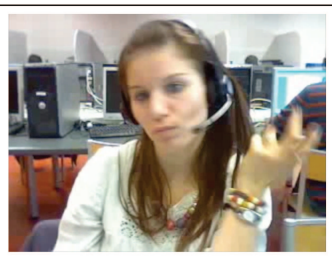
[1] Elle dit que BCBG est un « homme qui travaille », ce qui est un contresens car BCBG veut dire « bon chic bon genre ».

discours d'une française qui repère un bobo. Ce geste permet à Adèle de délivrer un discours rapporté et de changer ainsi de posture énonciative : elle passe d'une enseignante à une française qu'elle met en scène pour ses apprenantes.

9.2.2.5 Une focalisation multimodale sur le sens

Mélissa utilise à plusieurs reprises des expressions faciales pour se focaliser sur le sens de l'item lexical.

Transcription de la parole	**Image du participant**	**Phénomènes observés**
3 MEL en [fait ça]& 4 ALJ [non] 5 MEL &veut dire bourgeois bohème/ (0,9) j`vais t'envoyer une photo/ (1,7) c'est des gens un **peu chic** **(1,1)** [d'accord/] 6 ALJ [AH:\]		Regard caméra, yeux plissés, geste emblématique voulant dire « chic »
7 MEL mais:/ (.) un peu **cool aussi** (1,2) 8 ALJ [ce mot es^très:]		Yeux plissés, tête tournée, mimique faciale voulant dire « cool »
9 MEL [ils sont **chic/ (.)**		Chic
et cool] **(1,0)** 10 ALJ ce mot en espagne/ c'est très: mal\ ((rires)) (0,7) 11 MEL oui en france/ c'est un p`tit 12 ALJ ((rire))		Cool
13 MEL c'est **devenu un petit peu péjoratif** (0,9) m[ais euh\] 14 ALJ [ah:\] °d'accord° (0,3) 15 MEL c'est pas\		Regard caméra

	((13 secondes de réglages de problèmes techniques avec Ana))		
16 MEL	alors\ du coup/ hum: (1,6) est-ce que tu as donc euh: t'as compris/ alejandra c`que: (0,4) [c`que j'a]ttends&		
17 ALJ	[oui oui]		
18 MEL	&de toi/ (0,5) ouais/		
19 ALJ	oui		
	(1,4)		
20 MEL	donc hum: (1,5) alors\ (.) hum: (0,5) j`vais écrire un p`tit peu la clientèle que j`vise/ bon ba j'ai dit:/ des: des gens bobo:/ donc c'est des gens qui sont chic (1,5) et hum: (0,4) mais cool/ **(1,0)**		Expression faciale voulant dire « cool »
	((56 secondes d'interactions avec Alejandra))		
21 MEL	donc vous m- vous m'avez demandé: ce que je veux vendre et à qui/ (1,3) donc je voudrais vendre/ (0,3) à des/ (0,7) j- ma clientèle/ au quartier de croix-rousse/ (.) ce sont des bobos/ (1,2) d'accord/ (0,9) donc en fait j'ai envoyé une photo\ est-ce que tu vois la [photo/]		
22 ANA	[oui]		
23 MEL	((rire)) (0,4) ouais/		
24 ANA	bobo/ [c'est]:\ (1,3)		
25 MEL	[alors\]		
26 ANA	c'est [de l'arg]ot/		
27 MEL	[alors en fait/] (1,8) hu:m: (0,4)		
	non: c'est un mot nouveau:/		
	mais c'est **une contraction de**		Geste métaphorique qui tient le concept de « contraction »
	bourgeois/ (0,3) [bohème\]		Regard caméra
28 ANA	[ah:] okay		
	(1,5)		
29 MEL	bobo\ (0,2) **bourgeois bohème** (0,6)		Battement

	et donc ce sont des gens: **chic/ (0,3)**		Regard caméra
30 ANA	**[mais c]ool** [oui]		Regard caméra et mimique faciale voulant dire « cool »
31 MEL	**(0,6)** **d'accord/** (0,7) donc c'est cette clientèle que je vise\		Mimique faciale voulant dire « cool »

Transcription 14 : Une focalisation multimodale sur le sens

Mélissa explique l'item lexical deux fois : une fois pour Alejandra, quand Ana est absente à cause d'une panne technique, et une fois pour Ana quand elle revient. Mélissa aborde toutes les facettes de la connaissance de l'item lexical. Elle commence par décomposer le mot-valise uniquement à l'oral (ligne 5), contrairement à ses collègues qui utilisent aussi le clavardage. Quand elle réexplique la forme du mot-valise à Ana, elle utilise un geste métaphorique qui est redondant à la parole pour indiquer que c'est une contraction (ligne 27). Mélissa se focalise sur le sens en donnant une définition très simple « chic et cool », contrairement à Victor, qui donne une définition complexe et à Adèle, qui emploie une traduction, et à Samia, qui ne donne aucune définition. A la suite de cette remarque, Mélissa aborde enfin l'usage de l'item lexical pour chaque apprenante. À une remarque d'Alejandra (ligne 10), Mélissa montre que le mot « bobo » est un peu péjoratif (ligne 13), et à la suite d'une question d'Ana (lignes 24-26), elle parle du registre de la langue de l'item lexical en disant que ce mot est de l'argot (ligne 26).

En ce qui concerne l'expressivité de l'explication, ce qui est frappant c'est l'association de gestes et d'expressions faciales aux mots « chic » et « cool ». Quand Mélissa donne sa première définition, elle se touche les épaules, tourne sa tête et plisse ses yeux pour mimer l'adjectif « chic » (ligne 5). Elle maintient ce geste pendant une seconde avant de

baisser ses mains, tourner sa tête dans l'autre direction et plisser ses yeux à nouveau afin de mimer l'adjectif « cool » (ligne 7). Puis, elle répète les mots « chic » et « cool » et refait les mêmes mouvements (ligne 9). Après avoir abordé le registre, réglé un problème de connexion et vérifié la compréhension de la consigne de l'activité, Mélissa répète une troisième fois les mots « chic » et « cool » avec la même expression faciale voulant dire « cool » (ligne 20). L'expression faciale que fait Mélissa à la ligne 20 montre que les bobos sont selon elle prétentieux. Quand Mélissa explique le sens de l'item lexical à Ana, Mélissa répète pour une quatrième fois sa définition en reprenant la même expression faciale avec les mêmes mots (lignes 29-31). Cette redondance entre le geste et la parole fait en sorte que les apprenantes soient exposées au même contenu sémantique deux fois : une fois oralement et une fois visuellement. La répétition du même geste associé aux mêmes mots est probablement propice à la mémorisation.

Pour résumer, trois apprentis-enseignants sur quatre abordent toutes les facettes de la connaissance de l'item lexical « bobo » : sa forme, son sens et son usage. Tous abordent les composants du mot-valise, à la fois à l'oral et à l'écrit, sauf Mélissa qui ne les aborde qu'à l'oral. Tous sauf Samia abordent l'usage de l'item lexical, en parlant de sa fréquence et/ou du registre de langue dont il relève. A l'exception de l'envoi d'une photo, tous sauf Samia donnent une définition du sens de l'item. Ce constat nous montre qu'il n'est pas évident d'expliquer un tel item. En effet, cet item lexical est multiforme, et montre que quatre apprentis-enseignants qui tentent d'expliquer le même item lexical n'abordent pas les mêmes facettes de la connaissance de celui-ci.

On assiste à différents degrés d'expressivité pour expliquer l'item lexical. Victor et Samia utilisent uniquement le microphone et le clavardage, même si Victor marque clairement une transition de phases par la webcam. Ces deux apprentis-enseignants donnent des explications multimodales de la forme de l'item lexical et expliquent oralement et par écrit les composants du mot-valise. Victor utilise uniquement le microphone pour donner une explication lexicographique sans expressivité visuelle. Son explication alambiquée et peu expressive se révèle mal adaptée au niveau de son apprenant.

Adèle et Mélissa donnent des expressions nettement plus expressives. Adèle utilise des gestes pour signifier une traduction, pour changer de posture énonciative quand elle joue le rôle d'une française, et pour mettre à plat les composants du mot-valise. Mélissa associe les mêmes expressions faciales aux mots « chic » et « cool » quand elle explique le sens du mot « bobo », répétant quatre fois les mêmes mots et les mêmes mimiques faciales. Ses

apprenantes disposent ainsi d'une représentation visuelle susceptible de faciliter la mémorisation du terme.

9.2.3 Efficacité

L'efficacité des explications est difficile à évaluer. L'apprentissage du lexique n'étant pas un objectif explicite de la formation, nous n'avons pas administré des post-tests de vocabulaire aux apprenants, ni tout de suite après les séances, ni à la fin du semestre. Nous ne pouvons donc dire avec certitude si un item lexical a été acquis ou non. Tout ce dont nous disposons sont des indices d'uptake des apprenants, du type « oui j'ai compris ». Or, même ces indices d'uptake ne garantissent pas l'acquisition. Smith (2005) a étudié la négociation du sens par clavardage et a trouvé que la présence ou absence d'uptake pendant des activités en ligne ne semble pas avoir d'effet sur l'acquisition du lexique.

Cependant, nous pouvons définir certains critères d'efficacité, comme par exemple le temps passé sur l'explication et la négociation. Concernant le temps, nous pouvons mesurer le temps pris pour conduire une explication lexicale, du moment où le besoin est identifié par l'enseignant ou exprimé par l'apprenant jusqu'au moment où l'explication est terminée. Une séquence d'explication lexicale peu chronophage laisse plus de temps aux participants pour compléter l'activité. Concernant la négociation, on peut dire qu'une séquence dans laquelle il y a négociation est efficace pour au moins deux raisons. La première raison concerne l'objectif du semestre : s'entraîner à l'oral. Si les apprenants arrivent à parler et jouer un rôle actif dans l'interaction, l'objectif est réussi. Ensuite, comme nous l'avons évoqué dans le chapitre 3, les mots qui sont négociés sont susceptibles d'être mieux appris que les mots qui ne le sont pas (Ellis, Tanaka, & Yamazaki, 1994; Newton, 2013). Le temps passé sur l'explication et l'évaluation sont alors évalués.

Le tableau ci-dessous indique le temps passé sur l'explication par les quatre apprentis-enseignants, en ordre croissant :

Apprenti-enseignant	Samia	Mélissa	Adèle	Victor
Temps	50,7 s	52,3 s	78,6 s	119,5 s

Figure 2 : Le temps pris pour chaque explication du mot « bobo »

L'explication de Samia est la plus courte, mais elle est aussi la moins complète. Comme évoqué dans la section précédente, Samia n'aborde ni le sens, ni l'usage de l'item lexical.

L'explication de Mélissa est presque aussi courte, bien que celle-ci explique deux fois, une fois pour chacune de ses apprenantes. Elle arrive à combiner de manière succincte une explication orale très simple à une mimique faciale afin de résoudre rapidement la compréhension du terme.

Adèle prend plus de temps, car elle se focalise plus sur la composition du mot-valise, ce qui prend plus de temps que l'explication succincte de Mélissa. Elle anime l'usage de l'item quand elle aborde la fréquence et l'utilité, au lieu de mentionner simplement le mot « péjoratif » comme le font les autres. Deux autres facteurs prolongent le temps d'explication : un décalage technique important s'observe, ce qui fait que la communication est ralentie. Les deux participants doivent en effet attendre plusieurs secondes après chaque tour de parole avant la réception du message par l'interlocutrice. La seconde raison est qu'Alannah pose des questions, ce qui nécessite plus d'explications de la part d'Adèle.

Enfin, l'explication de Victor est de loin la plus longue. Ses deux explications lexicographiques ne parviennent pas à résoudre rapidement la compréhension du terme. Au contraire, elles présentent une difficulté de compréhension, comme en témoignent les nombreuses pauses qui parsèment le discours de l'apprenant, et sa compréhension qui n'est que partielle.

En consultant le débriefing qui a eu lieu le lendemain, nous avons retrouvé l'échange suivant :

Interlocuteur	**Transcription de la parole**
Victor	Moi j'ai dû expliquer, et au début j'ai parlé de classes sociales, et je crois que ça lui a pas parlé. Et à un moment, il m'a demandé « Mais qu'est-ce que c'est un bobo ? » Là j'ai dit ben c'est une personne au moment que j'ai dit que c'était un type de personne là il a là ça l'a éclairé.
Formateur	Quelque chose qu'il connaissait déjà *a priori* quoi en fait euh ?
Victor	Ben au moins ils avaient oui il avait une bonne idée oui c'est des gens il m'a dit c'est des gens qui aiment bien s'habiller j'ai dit oui oui c'est ça c'est des gens qui… Non mais il y a eu un souci…
Etienne	Ils font attention à leur apparence.
Victor	Oui c'est ça.

Transcription : 15 : Discussion des bobos lors du débriefing

On constate que Victor se rend compte que le début de son explication lexicographique n'a pas bien fonctionné. Pour Victor, la demande de reformulation de Liam arrive au hasard comme une surprise, inattendue. L'apprenti-enseignant croit aussi que la seconde partie de son explication, qui est quasiment pareille à la première, a réussi à résoudre

l'incompréhension. Victor semble penser que le constat que les bobos s'habillent bien venait de Liam, mais en réalité il vient de Victor. Liam dit que les bobos « ont des euh beaucoup de choses euh beaux et et jolies » (ligne 16). La capacité à vérifier l'efficacité de sa propre explication semble donc être une compétence à acquérir.

Se pose alors la question suivante : une explication plus courte est-elle toujours plus efficace ? L'objectif étant la réalisation de la tâche, on peut argumenter qu'une explication courte constitue une moindre diversion. En même temps, la discussion d'un mot à forte charge culturelle partagée (Galisson, 1991) n'est pas du temps perdu, surtout pour des apprenants qui vont devoir séjourner en France.

Dans le cas de Mélissa, une explication courte est aussi une explication efficace, car son explication succincte réussit à résoudre la compréhension. Pour Samia, on ne peut en dire autant. Celle-ci réussit à inciter l'apprenante à effectuer une recherche sur Internet, ce qui favorise l'autonomie, mais l'explication elle-même ne semble pas être très efficace. La séquence de Victor, comme celle de Mélissa, est un cas où s'observe une relation négative entre temps d'explication et efficacité. Son explication verbeuse ne parvient pas à résoudre efficacement l'incompréhension. Pour Adèle, elle prend plus de temps que Mélissa, mais elle aborde de manière plus expressive l'usage du terme. Le temps qu'elle consacre à l'explication semble alors rentabilisé.

9.2.4 Synthèse : bobo

L'analyse du caractère opportun ou inopportun des explications nous a permis de dégager certaines compétences nécessaires pour l'enseignement du vocabulaire en ligne. Si une explication lexicale est planifiée, comme c'était le cas pour les bobos, une réflexion sur le meilleur moment de l'introduire s'impose. Si l'opportunité d'introduire l'explication comme la réponse à une question de l'apprenant se présente, il faut la saisir, comme le font Mélissa et Adèle. Sinon il revient à l'apprenti-enseignant de trouver un moyen de réorienter la conversation vers une discussion du mot-clé de façon naturelle et de façon à ce que l'apprenant en comprenne l'intérêt. Nous avons observé cette réorientation stratégique chez Victor. Faute de cette réorientation, le terme-clé ne peut qu'être introduit inopportunément, et à l'emporte-pièce, ce qui risque d'avoir comme conséquence que l'apprenant ait du mal à comprendre l'intérêt du terme. L'introduction de Samia est l'exemple frappant.

L'analyse de la focalisation sur les différentes facettes de la connaissance d'un item lexical a révélé qu'expliquer un mot n'est pas isomorphe. Avant d'expliquer un mot inconnu,

l'apprenti-enseignant doit se poser la question : quelles facettes de la connaissance du mot sont nécessaires ? Tous les apprentis-enseignants se sont focalisés sur la forme. Tous sauf Samia se sont focalisés sur le sens (mais elle a envoyé une photo), et tous sauf Samia se sont focalisés sur l'usage. Il y a même des différences d'explication de la même facette : alors que Victor et Mélissa abordent le registre du mot en disant que le mot est péjoratif, Adèle parle de la fréquence du mot et non pas du registre. De même, concernant le sens, Victor donne une définition lexicographique qui parle de la classe sociale et des orientations politiques des bobos, tandis que Mélissa donne une description physique simple, et Adèle donne une traduction et une définition simple. Une fois les facettes sélectionnées, une réflexion s'impose sur la meilleure façon de les expliquer.

L'analyse de l'expressivité des explications nous a permis de dégager certains profils multimodaux. Concernant la focalisation sur la forme, Victor et Samia utilisent la voix et la clavardage pour expliquer les composants du mot-valise. Mélissa utilise la voix et un geste métaphorique qui indique que c'est une contraction de deux mots. Adèle utilise trois modes pour se focaliser sur la forme : sa voix, le clavardage, et tout son corps pour animer la décomposition du mot-valise en deux syllabes. Il semblerait donc qu'Adèle éprouve plus de *flexibilité multimodale* pour se focaliser sur la forme de l'item lexical. Concernant la focalisation sur le sens, Victor n'utilise que sa voix, et son corps reste figé. Mélissa et Adèle utilisent leurs corps pour mimer quelqu'un qui est « chic » et « cool » ou pour indiquer que le mot « hipster » est une traduction. Enfin, concernant la focalisation sur l'usage de l'item lexical, Victor n'utilise que sa voix pour dire que le mot est péjoratif. Mélissa accompagne la même remarque d'un regard vers la caméra, ce qui peut capter l'attention de son apprenante. Adèle anime l'usage en faisant un geste métaphorique qui l'aide à insister sur la fréquence d'usage, et elle produit un geste emblématique qui l'aide à changer de posture énonciative pour mettre en scène un usage typique du mot.

9.3 Goûter d'anniversaire

Dans cette section, nous analyserons les cinq séquences d'explication lexicale qui portent sur l'item lexical « goûter d'anniversaire ». Au lieu d'organiser les analyses par critères, nous proposons des analyses holistiques sur chaque séquence d'explication. La consigne de l'activité donnée par les apprentis-enseignants était : « On va faire un petit jeu de rôle. Tu es équipier chez McDonald's et je suis ton manager. Ensemble on réfléchit à une nouvelle formule pour les goûters d'anniversaire et on va construire un pense-bête, c'est-à-

dire une liste de toutes les actions à entreprendre pour mener à bien ce projet. Alors, d'après toi, qu'est-ce qu'il faudrait faire ? »[2] Ces analyses mettront en relief les différentes façons d'expliquer et d'impliquer l'apprenant, tout en tirant profit des affordances.

9.3.1 Séverine : une revue du lexique

Séverine semble bien consciente de l'importance de comprendre tout le vocabulaire de la consigne avant de se lancer dans la tâche. Elle demande à son apprenante de lire la consigne de l'activité à haute voix avant de lui demander s'il y a des mots inconnus.

Transcription de la parole		
1	SEV	je voudrais que tu lises/ (0,9) ce- (.) ce ce qu'il y a marqué sur le tchat\ (.) d'accord/ (0,7) vas-y\ (0,5) alors\
2	NAO	((l'apprenante lit les consignes à haute voix))
3	SEV	d'accord\ (0,2) alors\ est ce qu'il y'a des mots/ (0,3) que tu ne comprends pas\ (1,2) ou ça va\ (1,5)
4	NAO	mh::: (0,6) pense bête/ (1,0)
5	SEV	un pense bête/ c'est un: mémo (0,5) tu sais/ (.) c'est [une li]ste/ (.) okay/ (0,5) al[ors]\&
6	NAO	[OH] (1,7) [oui]
7	SEV	&0,4) tu vas/ (.) ah donc (0,3) euh tu as compris/ donc\ (0,4) est ce que tu sais ce que c'est équipier (0,5) chez mcdonald's (0,4) équipier/ c'est les salariés\ (0,4) d'accord/ (0,5)
8	NAO	oui (0,9)
9	SEV	okay/ (0,3) alors\ (0,5) euh goûter d'anniversaire/ tu comprends/ (0,3) mouais c'est [fac-] (.) c'- "birthday party"/
10	NAO	[mh mh]/

Transcription 16 : Une vérification de la compréhension de plusieurs items lexicaux

9.3.1.1 Un espace ménagé pour un travail lexical

Séverine passe d'une focalisation sur la consigne à une focalisation sur la langue. Ainsi se met-elle en mode explicatif. En demandant à Naomi « est-ce qu'il y a des mots que tu ne comprends pas ? » (ligne 3), Séverine ménage un espace de travail sur le lexique dans lequel Naomi peut mentionner tous les items lexicaux à comprendre. Naomi nomme le premier mot incompris, « pense bête » (ligne 4). Celui-ci est expliqué par deux synonymes (ligne 5). Puis, Séverine vérifie la compréhension des deux autres items lexicaux, « équipier » (ligne 7) et « goûter d'anniversaire » (ligne 9). Séverine semble avoir perdu de vue l'objectif de sa demande de la ligne 3 : au lieu de laisser la parole à Naomi pour détecter les mots incompris, Séverine recourt à des vérifications de compréhension et répond à celles-ci avant

[2] Nous ne disposons plus des messages exacts qui ont été envoyés par Visu. Nous reproduisons ici la consigne qui a été lue à haute voix par l'apprenante de Séverine.

toute réponse de Naomi. Séverine prive ainsi son apprenante d'opportunités de ressentir le besoin d'une explication, car le besoin est imposé par l'enseignante. L'explication est opportune dans le sens où elle arrive lors d'une discussion qui est explicitement lexicale, mais inopportune en même temps car l'apprenante n'en exprime aucun besoin.

9.3.1.2 Une complexité minimale

Séverine donne la forme orale et la forme écrite de l'item lexical. La forme écrite est comprise dans la consigne de l'activité et la forme orale est renforcée par l'usage de l'autonymie à la ligne 9. La focalisation sur le sens est incomplète car Séverine ne donne qu'une simple traduction. Elle n'aborde pas la coutume française de « goûter » et ignore sa charge culturelle.

9.3.1.3 Une faible expressivité

Séverine utilise uniquement sa voix pour expliquer le sens de l'item lexical. Elle n'utilise ni son corps, ni son visage, ni ses mains. L'image de Séverine n'apporte rien à l'explication. En effet, elle n'utilise pas la webcam à des fins communicatives. Ses yeux sont invisibles quand elle prend la parole au début de la ligne 1 :

Parole orale		**Image de Séverine**	**Phénomènes observés**
1 SEV	alors\ (0,5) euh goûter d'anniversaire tu comprends/ **(0,3) mouais c'est [fac-]&**		Séverine regarde l'image de son apprenante
2 NAO	[mh mh]/		
3 SEV	**&(.) c'- "birthday party"/ (0,2) alors**		

Transcription 17 : Une traduction rapide

Si cet item lexical était inconnu de Naomi, aucun indice multimodal n'augmenterait la mémorabilité.

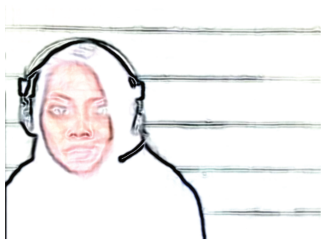
9.3.1.4 Une efficacité immensurable

Enfin, concernant l'efficacité, même si Séverine donne une explication avant que Naomi ne réponde à la vérification de compréhension, la réponse positive de Naomi à la ligne 2 indique que l'item lexical était connu et qu'une explication n'était pas nécessaire. Si l'item lexical avait été inconnu et si une explication avait été nécessaire, on aurait pu dire que la traduction donnée par Séverine a été partiellement efficace. Une telle traduction suffit pour instaurer une compréhension rapide afin de passer à l'activité, mais comme évoqué plus haut,

aucun élément culturel n'a été abordé. Aucune discussion sur les différences entre le goûter d'anniversaire en France et le goûter d'anniversaire en Irlande n'a eu lieu.

9.3.2 Etienne : une explication encadrée par l'activité

Etienne ne dévie pas de l'activité en expliquant l'item lexical :

Parole orale			Image d'Étienne	Phénomènes observés
1	ETN	euh: donc euh: (0,6) je sais pas si: vous connaissez c`principe **(1,7) hum:** d'anniversaire à mcdonald's\		Expression d'incertitude
2	SOH	**(1,4)** **°hum nan\°** (2,8)		Secouement de la tête pour indiquer son incompréhension. Son expression est semblable à celle d'Étienne ci-dessus.
3	ETN	**nan (0,4) ok** (0,5)		Secouement de la tête
		ben ici pa- c'est possible que: (0,2) un enfant/ (0,5) euh: invite tous ses amis au mcdonald's (.) et ensuite/ (1,0) euh VOUS vous devez vous occuper de l'organisation/ (0,4) de la **fête** (0,8)		Haussement de sourcils
4	ETN	ok/ **(1,1)**		Demi sourire
5	SOH	[ok]		
6	CNR	[oui]		

Transcription 18 : Une définition en lien avec la tâche (Etienne)

9.3.2.1 Une explication opportune avec un alignement socioaffectif

Etienne anticipe l'incompréhension du terme d'« anniversaire à McDonald's ». Contrairement à Séverine, Etienne attend une réponse à sa vérification de compréhension avant de passer au noyau de l'explication. Sa vérification de la compréhension est accompagnée d'une expression faciale qu'on peut qualifier d'une expression d'incertitude. Sophie comprend la vérification de compréhension et répond négativement avec un signe de la tête et une expression faciale similaire à celle d'Etienne. Etienne ratifie l'incompréhension en secouant sa tête. Puisqu'on observe des expressions faciales similaires, on peut parler d'un alignement socioaffectif.

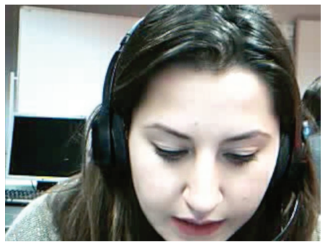
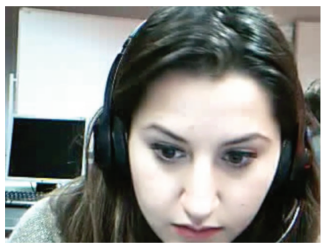
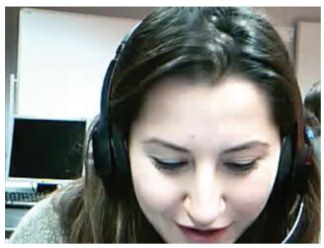
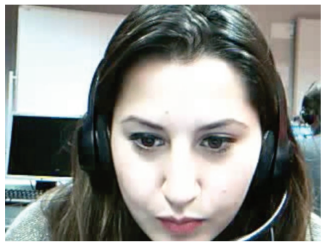
9.3.2.2 Une explication peu expressive orientée vers l'activité

Au lieu de demander aux apprenants d'imaginer le sens du concept, Etienne prend en charge lui-même le noyau explicatif. En un seul tour de parole, Etienne donne un exemple d'un anniversaire typique chez McDonald's et met en exergue la responsabilité qui est susceptible de tomber sur les apprenants lors de ce type d'évènement. Le noyau explicatif d'Etienne est orienté à la fois vers la tâche qu'auraient à accomplir les apprenants lors de l'activité en cours et vers des situations hypothétiques dans lesquelles ils sont susceptibles de se trouver lors de leur séjour en France (« c'est possible que »). Etienne adopte une conception minimaliste de la culture (Galisson, 1991) qui privilégie l'enseignement de savoirs dont les apprenants auront besoin pour la vie quotidienne. En soulignant le réalisme de cette tâche et la forte probabilité que cette responsabilité tombera sur eux, Etienne donne une raison pour apprendre ce concept. Or, l'explication est orientée vers la tâche plutôt que vers le concept d'« anniversaire à McDonald's » ou vers celui du « goûter d'anniversaire ».

L'explication est concise et efficace, puisqu'à la fin, les apprenants comprennent l'activité, malgré le fait que cette explication ne soit pas très expressive. Etienne hausse les sourcils lors de la prononciation du mot « fête » (ligne 3), mais il n'utilise pas vraiment d'autres ressources que le voco-verbal pour expliquer le sens de l'item lexical.

9.3.3 Samia : une explication simple

Samia anticipe une incompréhension possible et vérifie la compréhension. Elle s'attend à ce que le terme soit incompris car elle commence à écrire l'item lexical avant que les apprenants ne répondent à la vérification de compréhension :

Parole orale			Image de Samia	Phénomènes observés
1	SAM	donc ensemble vous allez réfléchir tous les deux/ (0,6) à une nouvelle formule pour les goûters d'anniversaire\ **est-ce que vous savez c`que c'est les goûters d'anniversaire\ (1,6)**		Samia regarde vers le clavier avant d'écrire
2	ANG	s'il y **a: beaucoup d'enfants qui viennent pour un anni**versaire/		Samia lève ses yeux quand Angela parle
3	SAM	**(0,5) voilà\ (.) v- les goûters d'anniversaire c'est ça c'est pour les a**nniversaires/ (0,8) des enfants en fait qui viennent fêter leur anniversaire (.) au mcdo\ <((clavardage)) goûter d'anniversaire>		Samia sourit suite à la confirmation de compréhension d'Angela
4 5	ANG SAM	**(3,0) au mcdonald's\ (1,2) oui (0,5)** d'accord/		Samia regarde les apprenants en attendant leur confirmation de compréhension

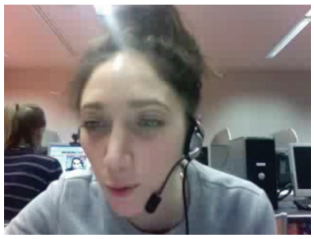
Transcription 19 : Une définition en lien avec la tâche (Emilie)

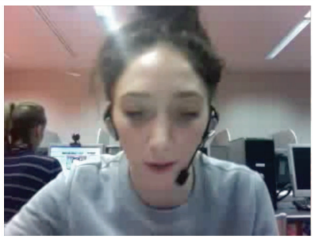
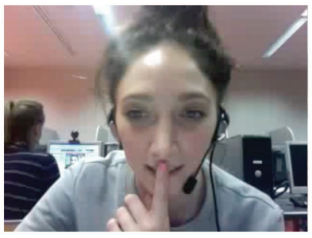
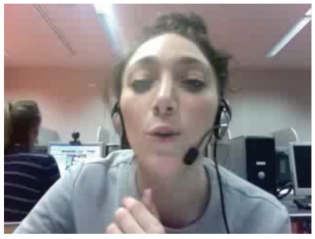
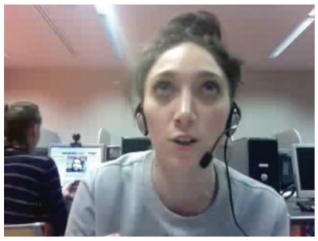
A la suite de la vérification de la compréhension, Angela prend en charge le noyau explicatif (ligne 2). Quand Angela commence à parler, Samia lève la tête, ce qui montre qu'elle est à l'écoute. Samia valide la définition d'Angela en souriant. Ensuite, elle ajoute des informations supplémentaires, ce qui vient corroborer les résultats de Smith (2003) et témoigne de l'utilité de son modèle dans lequel il y a une étape supplémentaire facultative nommée l'étape de « réaffirmation » qui consiste à donner davantage d'informations à la fin de la séquence. Samia reformule la définition d'un goûter d'anniversaire en l'orientant vers la tâche à l'instar d'Etienne et non vers la définition générale d'un goûter d'anniversaire.

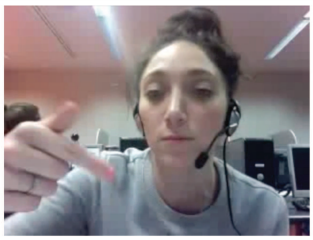
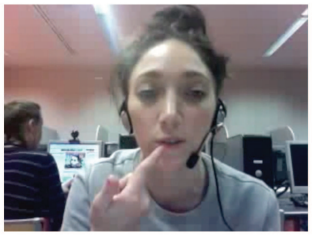
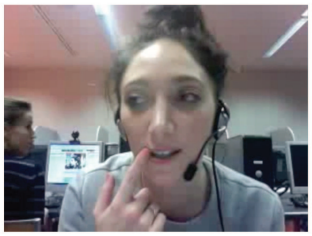
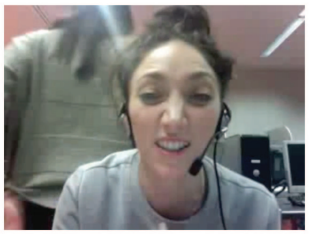
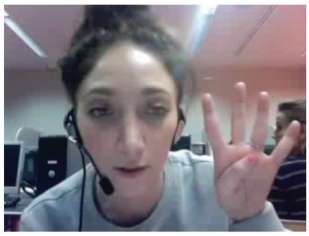
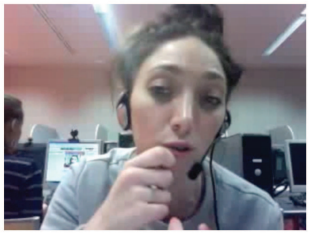
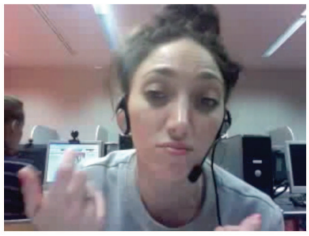
L'explication de Samia est opportune, car elle arrive après la vérification de la compréhension et après que l'apprenante donne une première définition. Elle aborde à la fois la forme orale et la forme écrite. Elle aborde plus la forme écrite que dans les explications précédentes, car celle-ci est plus présente grâce au clavardage (ligne 3). Le sens de l'item lexical est partiellement expliqué. Si son sens dans le cadre de l'activité est bien clair, aucun élément culturel n'est abordé. La définition de Samia n'est pas riche sémantiquement, d'autant plus qu'elle répète ce qui a déjà été dit par Angela. Concernant l'expressivité, le visage de Samia occupe tout le champ de la webcam, et celui-ci n'est pas très expressif mise à part quelques sourires. Avec Séverine, l'efficacité est aussi difficile à juger car Angela avait déjà une bonne connaissance de l'item lexical comme en témoigne la production orale de celle-ci à la ligne 2. L'explication de Samia est donc opportune, voire un peu complexe, et peu expressive.

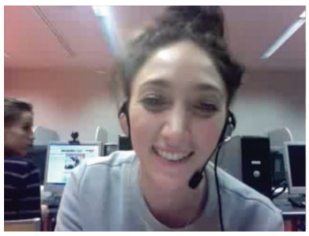
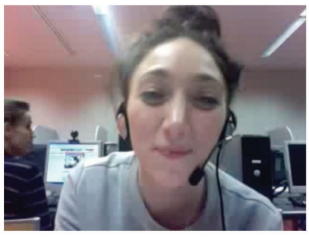
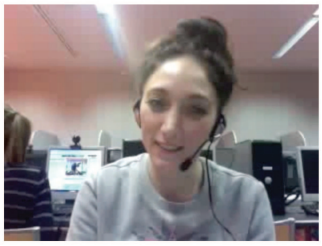
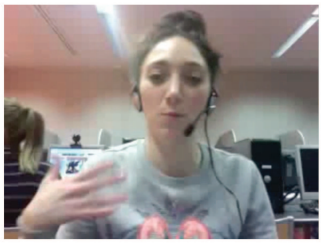
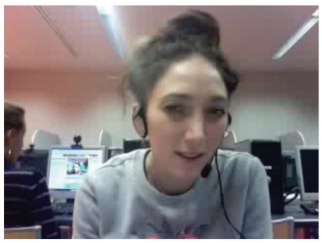
9.3.4 Adèle : une explication riche et expressive

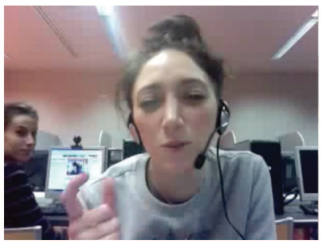
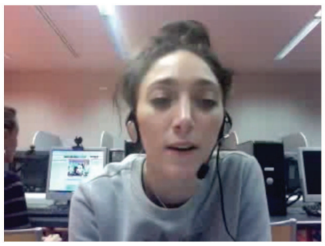
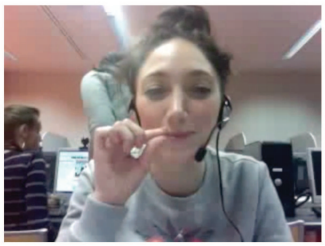
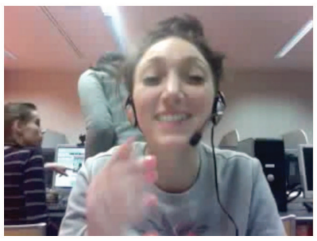
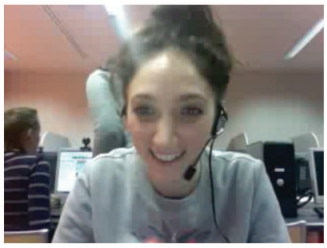
Quand Adèle explique les consignes de l'activité, elle ne vérifie pas la compréhension de l'item lexical « goûter d'anniversaire ». Elle préfère se focaliser sur le sens de la consigne plutôt que sur le vocabulaire de celle-ci. Alors qu'Adèle attend que les apprenantes réagissent à la consigne, Alannah demande tout de suite une explication du mot « goûter » :

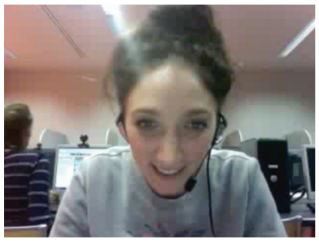
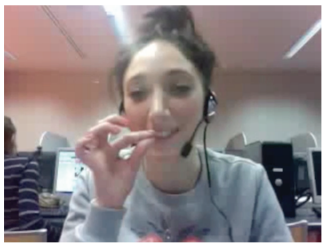
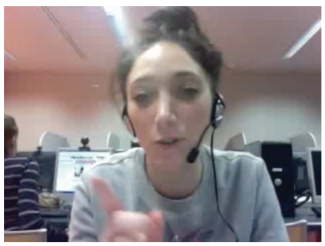
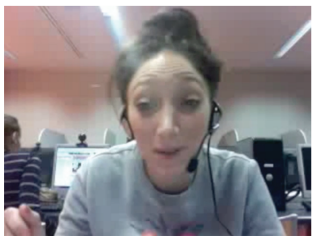
Parole orale		**Image d'Adèle**	**Phénomènes observés**
1 ALN	euh: **mh (0,4)** qu'est-ce que c'est euh les goûter "is it		Menton levé
2 ADL	**like the taste"/ (0,8) "or" euh [goûter/] (.) gouter d'a-**		Adèle se rapproche de l'écran
3 ALN	**["flav-"/]**		
4 ADL	**(0,4) euh [goûter/]**		
5 ALN	**["tas]te"/ (0,4) "yeah" (0,7)**		

6 ADL	alors (.) **(euh) goûter:/** (0,5) hop (.) attends\ (0,4) **ce mot là/** (0,9) <((clavardage)) goûters>		Deux regards vers le clavier lors de n'énonciation de « goûter » et « ce mot là »
	c'est euh: (0,2) **goûter d'anniversaire** comment est ce que je peux expli**quer goûter** (0,8)		Regard hors champ
7 ALN	**((rire))** **[un goûter:]** **["cause it's** like"]		Sourire
	(0,9)		Regard vers le bas
	je sais que c'est euh: (0,3) tas[te]/ mais\ <((clavardage)) taste>		Haussement de sourcils et hochement de tête
8 ADL	**[l-] (0,5) le** verbe le verbe goûter/ (.)		Regard pensif
	si tu **dis euh: (.)**		Regard pensif

	euh **je vais goûter/ (0,4)**		Index vers le bas, comme si elle allait récolter le glaçage d'un gâteau.
9 ALN	le **gâteau/** **ok (0,2) attends euh je** p- (.) je: euh (0,5) je rechercherai (0,3)		Fin de geste iconique pour goûter, hochement de tête « je p- »
10 ADL	un goûter pour les pour les en**fants/ c'est euh "a** birthday (.) birthday meal"\ (0,3)		Rapprochement de l'écran, regard pensif, haussement de sourcils lors de l'énonciation de « meal »
11 ALN 12 ADL	**"cake"** (0,2) "al:[right: okay"] [vous savez les en]f-		Expression d'incertitude
13 ALN 14 ADL	les enfants/ (.) peut-être ils font en irlande aus[si quand il est euh (.) euh **quat]re heures/** [°"four o'clock" (.) **x:°]** **(1,0)** m- milieu de la journée/ (.) [euh] (0,4)		Geste emblématique
15 CTR 16 ALN 17 ADL	**[on mange un:&** **[(inaud.)]** **[et euh je sais euh]** **&un un] un truc**		Geste iconique
	(2,3)		Geste emblématique pour dire « je ne sais pas » / « c'est une suggestion »

	tu sais/ **(2,2) tu expliqueras à catriona/**		Sourire
18 ALN	[16 secondes de silences et énoncés inaudibles] euh oui oui oui: **hu:m** (0,5)		Hochement de tête au début de la prise de parole d'Alannah
19 ADL 20 ALN	**["yeah" euh:]** **[((rire))]** **euh chez mcdonald's non non** non (0,4) euh: (0,6)		Rire, geste emblématique pour rire et éloignement de l'écran
	euh je j'ai bien ah compris maintenant/ mais euh je ne suis pas sûre/ **euh de ce mot/** **(0,8) mais euh (0,5)** je sais		Tête inclinée
21 ADL 22 ALN	**que goû- e**[uh goû- euh goûter/ euh] "is" euh [alors\ essaie de m'expliquer en français] (0,2) **[comme]nt tu vois** (inaud.) (1,5)		Hochements de tête. Rapprochement de l'écran et geste métaphorique
23 ADL	hum: "okay well" **(0,7)** **il f[aut** qu'on euh:] (.) x [tu vois c'est quoi le x] (0,6)		Tête inclinée pour montrer qu'elle est à l'écoute

	il n'y a pas **de synonyme** (1,8) 24 ALN "yeah" e[uh (.) euh je]& 25 ADL [c'est très dur] 26 ALN &comprends xx ah: (1,6)		Geste métaphorique commentant ses stratégies d'explication
27 ADL 28 ALN 29 ADL 30 ALN 31 CTR	[tu comprends/ (.) **et] toi&** [("well"; x:)] **&catriona** **pré- euh:** **°xx° quoi °x°**		Regard vers Catriona pour demander une réponse
32	[Vérification du fonctionnement du canal audio pendant 43 secondes]		
33 ADL 34 CTR 35 ADL	alors\ (.) catriona toi tu comprends le mot **euh goûter/** **(1,3)** euh oui je pense que oui hum: (0,2) une peti:te chose (0,5) pour\ m[ange]r& [oui]		Haussement d'épaules suite à la question, éloignement de l'écran suite au « oui » de Catriona
36 CTR 37 ALN	& une [petit chose c'est oui\] **[non: non non (ah; "I")]**		Adèle ratifie la compréhension de Catriona par un haussement de sourcils, un « oui » oral et un geste iconique
38 CTR	**"a snack"/ (0,4)**		Hochement de tête et geste déictique à l'émission de « snack » pour ratifier la compréhension
39 ALN	["snack"\] [catriona/] c'est pas ça c'est pour euh c`pour c'est pour ton anniversaire pour les les enfants (1,8)		Adèle regarde Alannah quand elle l'écoute

40 CTR	hum (.) "well" quand j'étais^en france le goûter c'était hum "snack" (0,4) "I don't know"\] (0,6)		Adèle regarde Catriona quand elle l'écoute
41 ADL 42 ALN 43 CTR	**c'est ça (0,5) [c'est un sna]ck** **["yeah"\]** **["ye]ah"\ (0,3)** "snack" (0,4)		Geste iconique pour manger un *snack*
44 ADL	mais (0,5) mais ce **que dit alannah/** c'est vrai donc c'est pour les enfants/ (.) et à mcdonald's/ (0,5)		Geste déictique représentant les propos d'Alannah
   45 ALN 46 ADL 47 ALN	quand ils parlent de goûter [d'anniversaire tu sais c'est] ["kinda like the kids' parties"] (0,8) voi:la (0,4) <((voix chantante)) **"surprise [party]">** [ok\]		Expression faciale et corporelle de soulagement suite à la phrase en anglais d'Alannah. Mouvement de corps pour mimer l'acte de danser en disant « surprise party »

Transcription 20 : Une explication riche et expressive

9.3.4.1 Une explication opportune et complexe

Puisque l'explication d'Adèle suit une demande de l'apprenante, on peut dire qu'elle est opportune et qu'il y a un besoin. Cependant, une explication ne semble pas avoir été préparée en amont. Plusieurs signaux corporels (regards hors champ, doigt sur les lèvres) indiquent qu'Adèle réfléchit à une définition et qu'elle est en difficulté (lignes 6-7).

Adèle aborde la forme, le sens et les éléments culturels de l'item lexical. L'explication est donc complexe. La focalisation sur la forme arrive pendant l'ouverture quand l'item lexical est décontextualisé et devient un objet de discours. Adèle le répète plusieurs fois oralement (lignes 2-8), le qualifie de « mot » et le répète par clavardage (ligne 6).

Adèle se focalise sur le sens par des moyens différents. Elle anime le verbe « goûter » par un geste (ligne 8), donne une traduction anglaise oralement, (ligne 10) mime l'acte de

« manger un truc » (lignes 14, 17), fait deux gestes pour « snack » (ligne 37, 41), et ratifie la traduction anglaise d'Alannah en disant « surprise party » (ligne 47).

L'usage de l'item lexical est aussi abordé. Adèle commence par donner un lien collocationnel (Benigno, Grossmann, & Kraif, 2015 ; Polguère, 2004) qui consiste à donner des liens sémantiques fréquents. Les mots « goûter » et « gâteau » allant souvent de pair, la connaissance de ce lien collocationnel fait partie de la connaissance de l'item lexical « goûter ». Pour renforcer la perception et la compréhension de ce lien, Adèle utilise un geste qui mime l'acte de goûter un gâteau (ligne 8). Puis, Adèle donne des informations culturelles sur l'usage du mot « goûter » en France, en disant que les enfants mangent à quatre heures de l'après-midi (ligne 12-17). Ces informations sont renforcées par un geste emblématique qui représente « quatre heures », et d'un geste iconique qui mime l'acte de manger.

Adèle arrive à impliquer ses deux apprenantes dans l'explication. Après avoir donné sa première explication, Adèle demande à Alannah de l'expliquer à Catriona (ligne 17), ce qui met Alannah en position d'explicatrice. A la suite d'une panne technique, Adèle demande à Alannah de lui expliquer le sens (ligne 21). Puis, Adèle vérifie par deux fois la compréhension de Catriona (lignes 27-29, & 33). A la suite de ces vérifications de la compréhension, les deux apprenantes dialoguent entre elles. Catriona pense que c'est un snack (lignes 36-38 & 40), tandis qu'Alannah pense que c'est un anniversaire pour enfants (ligne 39). Nous savons que la négociation est susceptible de favoriser l'acquisition du lexique (Ellis, Tanaka, & Yamazaki, 1994 ; Newton, 2013 ; Smith, 2003).

9.3.4.2 Une explication expressive

Tout au long de la séquence d'explication, Adèle est très expressive. Puisque la focalisation sur la forme se fait oralement et par clavardage, les apprenantes entendent, lisent, et prononcent l'item lexical plusieurs fois, ce qui aide la rétention (Nation, 2013). Mais ce qui est remarquable, c'est l'utilisation de la webcam par Adèle. Celle-ci utilise beaucoup la webcam pour des fonctions pédagogiques très différentes. Les gestes que fait Adèle présentent un intérêt particulier. En effet, ceux-ci remplissent les trois fonctions définies par Tellier (2008a) : informer, animer et évaluer, ainsi que des fonctions méta-explicatives.

Adèle utilise des gestes d'information pour se focaliser sur le sens de l'item lexical. Elle mime d'abord le verbe « goûter » par un geste à la ligne 8, puis par un nouveau geste à la ligne 14. Elle mime le mot « snack » par un geste à la ligne 37, et à nouveau à la ligne 41. Enfin, avec tout son corps et une expression faciale enthousiaste, elle mime la traduction « surprise party » en dansant dans sa chaise. Ces explications multimodales rendent

facilement perceptible le sens de l'item lexical. Les gestes d'information servent aussi à se focaliser sur l'usage de l'item lexical. L'heure à laquelle le goûter a lieu en France est très importante, et ces informations sont renforcées par le geste. Après avoir prononcé les mots « quatre heures » (ligne 12), Adèle maintient son geste emblématique pendant plus d'une seconde, sans parler, afin de marquer son importance. Le geste aide à marquer les informations les plus importantes.

Les gestes d'Adèle servent aussi à animer l'échange. Quand elle met Alannah en position d'explicatrice, Adèle fait un geste métaphorique invitant Alannah à s'exprimer (ligne 21). Quand elle demande une explication à Catriona, elle pointe un changement de direction du regard (lignes 27-31) pour lui donner la parole. Puisque les changements de regard sont difficilement perceptibles en visioconférence, les mouvements d'Adèle aident à gérer les tours de parole.

On retrouve plusieurs gestes d'évaluation. Pour approuver l'explication de Catriona (lignes 34-43), Adèle fait trois gestes. Elle fait deux gestes iconiques représentant « un snack » (lignes 37, 41). Le premier, à la ligne 37, est produit sans paroles, ce qui permet à Adèle d'évaluer l'explication de Catriona sans l'interrompre. Au milieu de l'explication de Catriona, Adèle hoche la tête et produit un geste vers la caméra pour approuver son explication sans l'interrompre (ligne 38). A la fin de l'échange entre les deux apprenantes, Adèle ratifie explicitement l'explication de Catriona à l'aide d'un geste voulant dire « snack » (ligne 41), et ratifie également la contribution d'Alannah à l'aide d'un déictique vers l'endroit abstrait où se trouvent les propos d'Alannah (ligne 44). Les deux gestes voulant dire « snack » (lignes 37, 41) sont donc à la fois des gestes d'informations et des gestes d'évaluation. Enfin, le dernier mouvement que fait Adèle, en mimant « surprise party » (ligne 46), sert à approuver la traduction d'Alannah. Ce dernier geste est donc lui aussi un geste à la fois d'information et d'évaluation.

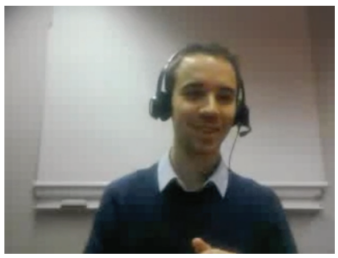
Les mouvements que fait Adèle devant la webcam ont aussi des fonctions méta-explicatives. Quand elle cherche la façon d'expliquer le mot « goûter d'anniversaire », Adèle met en scène sa réflexion en regardant hors champ, en souriant et en se touchant le menton (lignes 6-7). Le silence est ainsi comblé. A la suite de sa première définition, Adèle fait un geste emblématique pour dire « je ne sais pas, c'est une suggestion » (ligne 17). Ce geste montre son incertitude à l'égard de la suffisance de l'explication. Quand Alannah éprouve des difficultés à expliquer, Adèle mentionne qu'il n'y a pas de synonyme (ligne 23). Ce commentaire méta-explicatif est accompagné d'un geste métaphorique qui tient le concept de

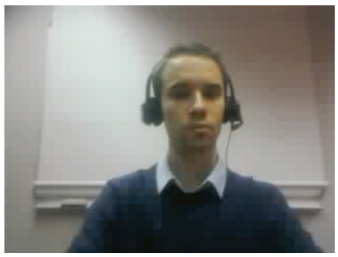
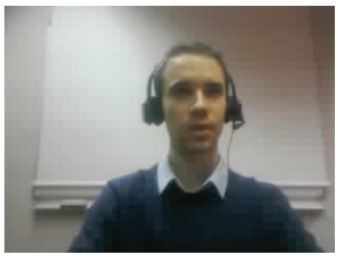
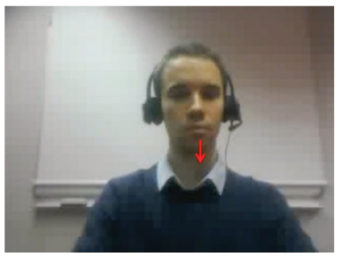
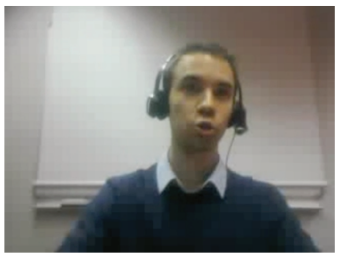
« synonyme » entre ses doigts. Le geste aide à cet accompagnement sociocognitif quand Adèle sympathise avec Alannah sur les difficultés à expliquer.

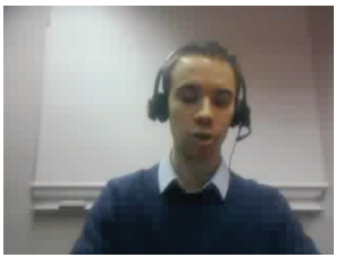
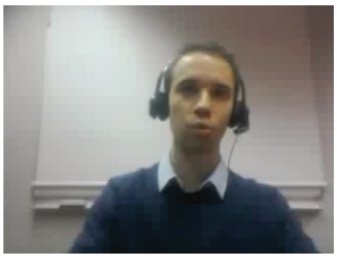
Bien que chronophage, surtout avec les deux pannes techniques au milieu, cette séquence d'explication est efficace car une multitude de facettes de la connaissance de l'item lexical a été expliquée avec expressivité. Les apprenantes ont eu aussi l'occasion de négocier entre elles le sens de l'item lexical. Cette séquence a été efficace et réussie, car elle a apporté une complexité, une expressivité, une opportunité, et une négociation entre apprenantes. Il y a eu une grande implication des apprenants et elles semblent avoir compris le sens de l'item lexical.

9.3.5 Victor : une séquence d'explication polygérée

Victor, comme Séverine, Etienne et Samia, s'intéresse au vocabulaire quand il explique la consigne de l'activité. Il vérifie la compréhension de l'item lexical « jeu de rôle » (ligne 1) et accepte la réponse incertaine de Liam sans le pousser à le définir. Puis, il vérifie la compréhension de « goûter d'anniversaire », déclenchant ainsi une séquence d'explication lexicale :

Parole orale	**Image de Victor**	**Phénomènes observés**
1 VIC donc on va faire un: petit jeu de rôle\ (.) tu sais c`que c'est un jeu de rôle/ (1,9) 2 LIM euh: (0,4) oui je je pense que euh: (1,0) je comprends (1,0) 3 VIC d'accord\ donc tu vas faire euh un employé chez **mcdonald's/ (1,0) et [moi je suis ton:] manager\&** 4 LIM **[((rire)) oui]** 5 VIC **&(0,4) d'accord/** (0,5) 6 LIM d'accord [d'accord]		Sourires
7 VIC [et ensem]ble en **fait euh on** va réfléchir à une euh nouvelle formule pour les: goûters d'anniversaires\ (0,5) est ce-que **tu sais** c`que c'est un goûter d'anniver**saire\ (1,5)**		Regard pensif

8 LIM euh: **(0,6) je pense euh (1,6) oui/ (0,7)** 9 VIC est-ce-que tu peux m'expliquer/ (.) c`que tu crois qu`c'est/ (1,4)		Froncement de sourcils, visage tendu
10 LIM euh: (1,1) un {gyt} **euh:** (0,7) .tsk (0,4) c'est (0,5) "ta[s]-" (0,2) ["taste"/] 11 VIC [j`]vais [t'aider j`vais] l'écrire/ hum\ (2,1) <((clavardage)) goûter d'anniversaire> hop (2,8)		Difficulté extrême
12 LIM **euh oui/ (0,3) hum: (0,7) en |anglais c'est euh (1,4) euh: (0,2) "taste of" (0,4) euh: (0,2) "a birthday"/**		Regard écran, visage peu expressif
(1,8) 13 VIC **d'accord/ (0,5) et qu'est-ce que c'est/** (1,4) 14 LIM hum: (1,1) c'est une: (0,8) euh: célébration/ (0,8) de: (0,2) votre (1,3) euh: âge/ (0,7)		Regard pensif, hochement de tête et inclinaison de tête avant d'écouter Liam
15 VIC **hm hm/ (1,0)** ok [très bien]\ 16 LIM [hu]m: (2,1) 17 VIC c'est cela\ (0,5)		Hochement de tête de Victor
donc tu sais je sais pas **moi je: je suis jamais allé au (.) au mcdo** pour faire ça/ mais euh\ (0,6)		Regard pensif

ils **euh: organisent des: des goûters/** (0,3) **euh: et comme ça euh: (0,5) la famille ne fait que réserver/**		Regard vers le bas
et ils préparent une petite **fête pour euh souligner l'anniversaire** d'un enfant\ (1,0)		Regard pensif
(0,5)		Hochement de tête pour indiquer sa compréhension

Transcription 21 : Une implication cognitive forte

Cette explication est très opportune car la compréhension de l'item lexical est nécessaire pour la réalisation de la tâche, et elle arrive après que Victor en vérifie la compréhension.

Victor aborde la forme orale et la forme écrite (ligne 11) avec le commentaire méta-sémiotique « hop » pour attirer l'attention vers la fenêtre de clavardage. Victor pousse Liam à définir le sens de l'item lexical (ligne 14) et Liam ajoute des informations culturelles (ligne 17) qui sont en lien avec l'activité. Les trois facettes de la connaissance sont alors abordées.

9.3.5.1 Expressivité moindre, implication forte

Hormis la focalisation multimodale sur la forme de l'item lexical, l'explication est peu expressive car Victor n'utilise pas la webcam. Il utilise uniquement sa voix. Le regard de Victor, qui change de direction entre le clavier, l'écran et le hors champ, ne semble pas contribuer à la séquence d'explication. Ce qui est remarquable, c'est la façon dont Victor implique son apprenant tout au long de la séquence (voir Figure 3 ci-dessous). Quand Liam répond affirmativement à la vérification de la compréhension par Victor, ce dernier lui demande une explication. Liam traduit littéralement, ce qui produit un jargon. Puis, Victor demande une reformulation. Enfin, Liam donne une définition acceptable en français. Victor

implique Liam en demandant des explications supplémentaires après la première réponse de celui-ci. Comme le note Fasel Lauzon (2014 : 123), « demander une explication après une réponse d'élève est un moyen pour l'enseignant de pousser l'élève à développer sa réponse ». Victor complète la définition, mais seulement après avoir *fait produire* l'apprenant. On assiste donc à une séquence d'explication réussie et efficace.

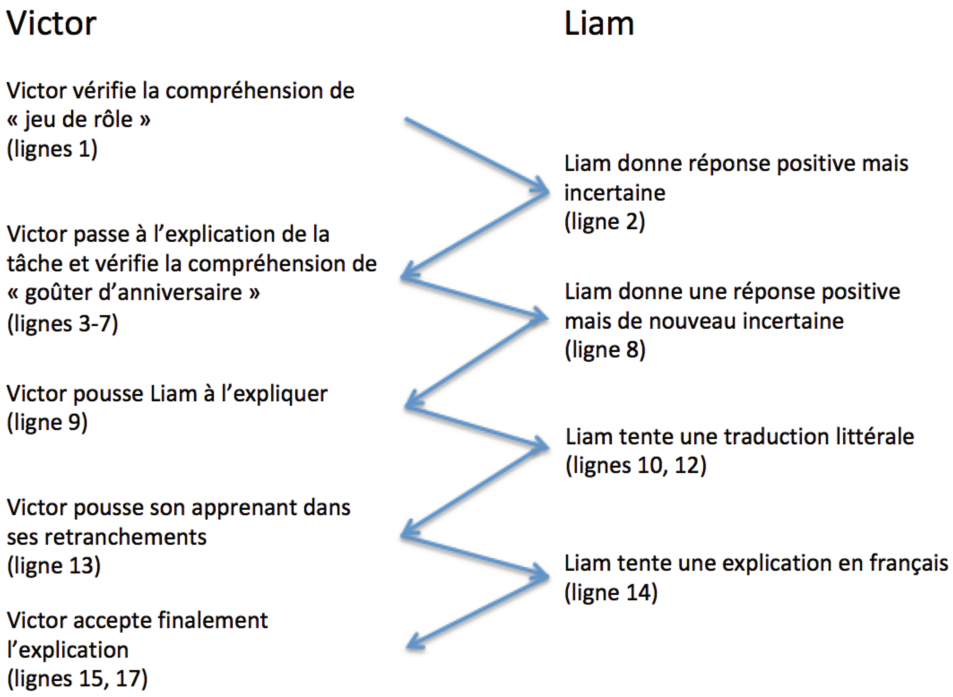


Figure 3 : Victor pousse implique son apprenant dans la séquence d'explication

9.3.6 Synthèse : goûter d'anniversaire

Ces analyses viennent corroborer les résultats des analyses précédentes sur les bobos. On constate qu'il y a une grande variabilité interindividuelle. Nos critères d'analyse, opportunité, complexité, expressivité et efficacité ont mis en relief ces différences. Toutes les séquences d'explication sur l'item « goûter d'anniversaire » ont été opportunes car elles arrivent au moment de l'explication de la consigne de l'activité. En outre, la connaissance de l'item lexical est nécessaire pour réaliser l'activité. Or, l'exemple de Séverine a montré qu'il est important de bien vérifier la compréhension de l'apprenant avant de passer au noyau explicatif pour que le besoin se fasse ressentir.

Seulement deux explications sur quatre ont été qualifiées de complexes. Si tous les apprentis-enseignants se sont focalisés sur les formes orale et écrite de l'item lexical, ainsi que sur le sens, la charge culturelle reste largement ignorée.[3] Adèle et Victor ont réussi à se

[3] Nous ne sommes pas en mesure d'expliquer pourquoi il y a cette différence.

focaliser sur cette charge culturelle. Adèle va beaucoup plus loin que Victor en parlant du rituel français qui est surnommé « le quatre heure », mais Victor parle tout de même de la possibilité qu'ont les Français de réserver un goûter d'anniversaire dans une chaîne de restauration rapide. Concernant la focalisation sur le sens, Etienne et Samia donnent des explications bien encadrées par l'activité, tandis que les autres donnent des définitions plus générales.

Adèle est la seule qui a été expressive, utilisant gestes, mimiques faciales, et regards pour animer la séquence d'explication. On pourrait dire donc qu'elle a fait montre d'une grande flexibilité multimodale. Les quatre autres n'ont pas vraiment utilisé la webcam et n'étaient pas particulièrement expressifs. Séverine est un cas déviant, car elle ne montre pas la totalité de sa tête.

Toutes les explications sont efficaces car tous les apprenants ont réussi à comprendre l'activité. Or, pour Séverine et Samia, dont les apprenants connaissaient déjà le sens de l'item lexical, un jugement sur l'efficacité demeure hasardeux.

L'analyse de ces cinq séquences d'explication nous donne aussi l'occasion de revisiter une stratégie d'explication que nous avons déjà vue dans le chapitre 7 : la traduction. Des traductions d'anglais en français ou l'inverse sont présentes dans les séquences de Séverine, Liam et Adèle. Dans chacune de ces séquences, la traduction est utilisée différemment. Séverine utilise une traduction immédiatement comme première stratégie d'explication pour vérifier la compréhension. Liam traduit littéralement « goûter d'anniversaire » en disant « taste of a birthday », ce qui est un non-sens. Adèle utilise une traduction à la fin de sa séquence pour donner raison à Alannah, qui est en désaccord avec Catriona sur le sens du mot « goûter ». Le lendemain de la séance, Adèle s'est exprimée sur ce sujet :

Apprenti-enseignante	**Parole**
Adèle	Expliquer « goûter d'anniversaire ». Donc, ça c'était, euh... Je me suis retrouvée, vraiment, alors... « Euh, goûter ! ». Là elle me dit euh « oui c'est taste » ! Alors j'ai fait « non, ça c'est le verbe. Le verbe goûter ». Mais elle me dit « c'est taste ». Voilà. Donc là j'ai dit « non », et puis euh, donc, j'ai essayé d'expliquer j'ai dit bon, alors il y en a une, Catriona qui me dit « Ah, quand j'ai travaillé en France, c'est manger une petite chose ». J'ai dit « oui, c'est ça, à quatre heure par exemple c'est le goûter », et euh, Alannah qui fait « Non ! C'est pas ça ! C'est pour les enfants ! » C'est, et là, euh, elle essaie donc j'ai dit ben « explique à Catriona ». Donc elle lui explique en français un moment, puis je vois que Catriona elle était là « oui mais on mange une petite chose ». Et là j'ai fait alors lui dire oui mais goûter d'anniversaire au McDo elle lui fait « It's a kids' party ». Elle lui dit en anglais et l'autre tout de

	suit « ah oui d'accord ». Donc il y a souvent ce recours un peu à l'anglais que je trouve (inaudible) au moins une ou deux fois sur ça, j'ai pas en tête c'était plus il y a une ou deux séance où j'avais noté ça. Euh, c'est, je trouve que ça fait sens, et ça fonctionne vraiment, enfin. Ça leur remet tout de suite quelque chose en tête, enfin c'est une référence, c'est "ah!". Ah ok ce mot c'est ça c'est ce mot là que je connais c'est comme ça que...

Transcription 22 : Discussion du goûter d'anniversaire lors du débriefing (Adèle)

Adèle commence par remarquer que son apprenante ne comprenait pas que c'était un verbe. C'est peut-être la raison pour laquelle Adèle a fait un geste iconique quand elle a donné la collocation « goûter le gâteau » (ligne 8). Elle parle ensuite de son action conscientisée d'impliquer les deux apprenantes à la fois en initiant une séquence de négociation du sens entre les deux. Adèle parle du désaccord entre ses deux apprenantes, et remarque que la traduction peut servir à résoudre la différence d'avis, et donner une signification à laquelle elles peuvent relier un nouvel item lexical jusqu'alors inconnu.

9.4 Conclusion

Ces comparaisons qualitatives sur deux explications nous ont permis de dégager des différents profils d'enseignement, et des différents usages de la multimodalité. Ces descriptions qualitatives sont complétées par des descriptions quantitatives qui se trouvent en annexe 6 en page 389.

Victor privilégie surtout le mode verbal, donnant des définitions lexicographiques trop complexes pour son apprenant. Dans les deux exemples, il n'utilise pas la webcam pour expliquer, mais il l'utilise pour marquer un changement de phases. Il utilise beaucoup le clavardage pour pointer les formes des items lexicaux. Il fait en sorte que son apprenant participe à la séquence d'explication du goûter d'anniversaire, mais ne réussit pas à faire participer activement son apprenant lors de l'explication des bobos.

Samia préfère des explications très simples. Son explication sur les bobos porte uniquement sur la forme de l'item lexical, alors que son explication sur le goûter porte uniquement sur l'activité. Comme Victor, elle utilise le clavardage pour pointer la forme des items lexicaux et bien que son visage soit visible, on ne peut pas la qualifier d'expressive car elle n'utilise pas vraiment la webcam pour gérer les séquences d'explication.

Adèle est plus expressive que les autres. Dans les deux séquences d'explication observées dans ce chapitre, elle utilise son corps, ses mains, et son visage pour se focaliser sur la forme, sur le sens, et sur l'usage des items lexicaux. Elle fait aussi des gestes méta-

explicatifs pour marquer le statut de certaines définitions et pour changer de posture énonciative.

On peut dire donc qu'Adèle montre une grande flexibilité intermodale et une grande flexibilité modale. Elle montre une flexibilité modale quand elle utilise un seul mode, par exemple les gestes, pour se focaliser sur la forme, sur le sens, sur l'usage, pour impliquer ses apprenantes, et pour changer de posture énonciative. Le mode *gestes* est flexible car il est capable de remplir toutes ces fonctions. Adèle a une flexibilité multimodale quand elle utilise plusieurs modes à la fois : sa voix, le clavardage, son visage, et ses gestes, pour remplir une seule fonction, par exemple se focaliser sur la forme du mot « bobo ». Peu d'études ont comparé les profils multimodaux des enseignants. Il pourrait être enrichissant d'étudier les concepts de flexibilité multimodale et de flexibilité modale à l'aide de données contrôlées.

Chapitre 10 : Discussion

Introduction

Les analyses des quatre chapitres précédents nous ont permis de répondre à nos questions de recherche et de développer certaines notions. Nous allons dans un premier temps résumer les principaux résultats, les notions développées, et leur utilité pour la formation des enseignants. Dans un second temps, nous allons aborder des limites de l'étude.

10.1 Synthèse des résultats et propositions didactiques

Nous nous sommes demandé, dans le premier chapitre, combien y avait-il de séquences d'explication lexicale dans les vingt-huit séances que nous avons transcrites. Nous avons repéré 295 séquences d'explication lexicale qui se répartissent presque également entre des problèmes de production et des problèmes de compréhension. 103 séquences sont déclenchées par l'apprenant, 133 sont déclenchées par l'enseignant, et 59 sont déclenchées par les deux à la fois. Ce résultat corrobore celui d'Inceoglu (2015) qui a trouvé que la plupart des séquences d'explication lexicale sont déclenchées par l'enseignant. Il y a en moyenne 10,5 séquences d'explication lexicale par séance. Ces résultats démontrent que, même si l'acquisition du vocabulaire n'est pas l'objectif premier du module de formation, les participants y consacrent du temps et de l'énergie. Que les séquences d'explication lexicale soient planifiées ou non, elles constituent un phénomène incontournable dès lors que l'on adopte une approche émique qui s'intéresse aux préoccupations des participants. Les chercheurs et les formateurs ont donc intérêt à se demander de quelle façon les séquences d'explication lexicale sont initiées et gérées par les participants pour ensuite réfléchir à des pistes pour améliorer la formation professionnelle des enseignants. Dans les paragraphes suivants, nous allons synthétiser les résultats obtenus lors de chaque chapitre d'analyse, tout en formulant des propositions didactiques.

10.1.1 Focalisation sur la forme

Dans le chapitre 6, nous avons analysé les fonctions pédagogiques remplies par la focalisation sur la forme des items lexicaux lors des problèmes de compréhension et de production. Les logonymes, dislocations, et renforcements par clavardage constituent les moyens verbaux les plus fréquemment employés par les apprentis-enseignants pour se

focaliser sur la forme. Les moyens visuels comme les expressions faciales, mouvements de sourcils, regards, et gestes, sont utilisés dans une séquence sur six pour se focaliser sur la forme. Nous ignorons si ces mouvements visuels sont conscientisés ou si les apprenants y prêtent attention, mais le marquage multimodal d'items lexicaux mérite d'être sujet de discussion lors des rétrospections et séances de débriefing en aval des interactions pour sensibiliser les enseignants aux différents effets qu'ils peuvent produire sur les apprenants.

Au cours des problèmes de compréhension, la focalisation sur la forme aide les participants à s'orienter mutuellement vers l'item lexical qui fait l'objet de la séquence d'explication. Cette focalisation est primordiale pour que la séquence d'explication puisse résoudre l'incompréhension. La focalisation sur la forme aide les apprentis-enseignants à repérer et problématiser l'élément non compris dans le cas où l'apprenant ne souligne pas l'élément non compris, ce qui est souvent le cas. Ce résultat corrobore ceux de Miecznikowski (2005). Pour les mots transparents ou phonétiquement proches, une focalisation sur la forme peut résoudre rapidement l'incompréhension. Le clavardage se prête particulièrement bien à cette focalisation car une difficulté de compréhension ou de production que peut avoir un apprenant à l'oral peut ne pas exister à l'écrit (Hilton, 2003). Enfin, la focalisation sur la forme permet à l'apprenti-enseignant de relier le sens d'un item lexical à sa forme.

Dans le cas d'un problème de production, la focalisation sur la forme permet aux apprentis-enseignants de :

- corriger la prononciation ;
- corriger l'orthographe ;
- étendre les connaissances productives des apprenants ;
- distinguer entre mots phonétiquement proches ;
- rendre perceptibles les mots justes qui sont donnés à la suite d'une panne lexicale.

Nous préconisons de souligner la forme des items lexicaux lorsque l'incompréhension de ceux-ci est anticipée ou manifestée. Cela permettra aux apprenants de les noter afin de les réviser en dehors des cours. Cela permettra également à l'ouverture de la séquence de « projete[r] l'explication comme action subséquente nécessaire » (Fasel Lauzon, 2014 : 223). Nous préconisons l'utilisation du clavardage lors de la mise en exergue d'un item lexical complètement inconnu pour que l'apprenant soit exposé à la fois à la forme orale et à la forme écrite. Nous insistons avec Codreanu & Combe Celik (2012) sur l'importance du clavardage

pour donner l'orthographe d'un mot. Le clavardage n'est pas un canal qui est utilisé en permanence, contrairement à la webcam qui est utilisé *a priori* tout le temps et au microphone qui est utilisé très souvent. Par conséquent, nous faisons l'hypothèse que son utilisation ponctuelle et stratégique est plus à même d'attirer l'attention de l'apprenant. Nous avons constaté une variation de l'utilisation du clavardage par les sept apprentis-enseignants. Adèle, Mélissa, Samia et Victor utilisent le clavardage plus que les autres pour se focaliser sur la forme des items lexicaux, ce qui peut avoir un impact positif sur l'efficacité des séquences d'explication. L'utilisation de cet outil mérite attention lors de la formation des enseignants.

Concernant la répétition d'un item lexical oralement et par écrit, nous avons développé la notion de **pléonasme multimodal** pour désigner la répétition d'un item lexical par ces deux modes différents. Bien que la redondance puisse nuire à l'apprentissage (Mayer, 2001; Sweller, 2005b), nous avons justifié la pertinence de pléonasmes multimodaux dans ce contexte particulier en raison du statut de non-expert des apprenants face à la langue, et en raison de la complexité de la connaissance d'un item lexical. En effet, le principe de redondance ne s'applique pas à tous les apprenants de la même manière. Ce qui est redondant pour un expert ne l'est pas nécessairement pour un novice. Puis, les formes orale et écrite d'un item lexical constituent deux parties distinctes de la connaissance de celui-ci et à ce titre ne sont pas redondantes. Nous sommes donc convaincu que les pléonasmes multimodaux ont leur place dans l'enseignement du vocabulaire par visioconférence.

10.1.2 Focalisation sur le sens

Nous avons analysé dans le chapitre 7 les stratégies mises en place par les apprentis-enseignants lorsqu'ils se focalisent sur le sens des items lexicaux. Nos analyses comprennent trois parties : le recours à la langue maternelle, les explications en français, et les explications ostensives. Les définitions, traductions, synonymes et exemples constituent les types d'explications verbales les plus employés, et les gestes sont utilisés dans une séquence d'explication sur six.

Les traductions sont très utilisées aussi bien par les apprentis-enseignants que par les apprenants. En effet, les traductions sont utilisées par les apprenants dans plus d'une séquence d'explication sur trois : elles ne peuvent donc être ignorées. Elles sont employées moins souvent par les apprentis-enseignants et se trouvent dans presqu'une séquence sur six. Elles servent à résoudre rapidement l'incompréhension à condition que l'apprenti-enseignant connaisse le mot dans la langue maternelle de l'apprenant. Nous avons constaté une

différence d'avis concernant l'emploi de traductions : les contrats didactiques que mettent en place certains apprentis-enseignants, souvent de manière implicite, considèrent les traductions comme répréhensibles, tandis que d'autres les considèrent comme une stratégie utile. Selon nous, les traductions sont acceptables quand il s'agit de vérifier la compréhension de l'apprenant, mais non pour les explications initiales. Au niveau intermédiaire, les apprenants devraient favoriser la création de liens entre éléments dans la L2 au lieu de relier des nouveaux éléments à des formes en L1. Les explications en français prépareront mieux les apprenants pour leurs stages en France l'année suivante.

Les synonymes, comme les traductions, peuvent résoudre rapidement l'incompréhension d'un item lexical et désambiguïser deux mots qui sont phonologiquement proches. Nous avons démontré que les synonymes et les définitions courtes semblent être plus efficaces que les définitions longues et complexes, ce qui corrobore le constat de Chaudron (1982) ainsi que celui de Mayer (2001) que les explications courtes et concises sont à privilégier. Le clavardage est très peu utilisé pour expliquer le sens des items lexicaux : taper moins vite qu'on ne parle pourrait en être une raison. Selon nous, le clavardage est mieux adapté à la focalisation sur la forme des items lexicaux qu'à la focalisation sur le sens. Les apprentis-enseignants semblent partager cet avis.

Nous avons démontré que les explications ostensives peuvent être données en utilisant tous les canaux de la webcam à l'exception du clavardage. Le microphone peut être utilisé pour donner une explication ostensive comme le fait Samia dans l'exemple 22 en page 245. Cette utilisation témoigne d'une certaine conscientisation de l'une des particularités de la plateforme : l'amplification des bruits de bouche par le microphone.

Les gestes et les expressions faciales sont utilisés par les apprentis-enseignants pour se focaliser sur le sens des items lexicaux dans une séquence sur six. Certains apprentis-enseignants comme Adèle utilisent leurs expressions faciales pour véhiculer du sens, ce qui témoigne du rôle important de celles-ci en visioconférence (Cosnier & Develotte, 2011). Nous n'avons pas mesuré les gestes qui ont été produits hors du champ de la webcam comme l'ont fait Guichon & Wigham (2016), mais parmi les gestes qui sont observables par les captures d'écran, plusieurs semblent être conscientisés, surtout ceux de Victor comme nous l'avons vu dans l'exemple 6 du chapitre 5 en page 160, et dans l'exemple 24 du chapitre 7 en page 248. Ces gestes peuvent être utilisés pour expliquer des concepts concrets et abstraits. Il convient toutefois de rappeler aux futurs enseignants que le nombre d'images par seconde peut être très faible à cause du débit de la connexion, ce qui a un impact sur la fluidité de l'image. Nous

préconisons donc que les formateurs rappellent aux futurs enseignants qu'il faut non seulement rendre les gestes visibles, en veillant au cadrage de la webcam, mais aussi ralentir les mouvements et maintenir les gestes plus longtemps devant la webcam à la suite de chaque mouvement.

Les apprenants semblent être conscients de l'utilité des gestes en visioconférence comme en témoigne l'échange entre Mélissa et Ana dans l'exemple 25 du chapitre 7 en page 251. Cet exemple a montré que des formulations provisoires (Lüdi, 1994) peuvent être produites par les apprenants non seulement par des moyens verbaux (cf. chapitre 7 exemple 10 en page 229) mais aussi par gestes. Les apprentis-enseignants ont donc intérêt à prendre conscience de cette réalité, et à surveiller la production gestuelle des apprenants.

Les images ne sont utilisées que rarement pour expliquer le sens des items lexicaux, et il semble que les apprentis-enseignants soient peu enclins à les employer. Nous avons vu dans le chapitre 7 (exemple 20 en page 240) que quand Samia envoie une image pour expliquer le sens du mot « quenelle », elle est incapable de donner une explication verbale. Elle aurait dû commenter l'image oralement tout de suite après l'avoir envoyée afin que les apprenants puissent comprendre le sens du mot et pour que les processeurs visuel et auditif (Mayer, 2001) soient exploités ensemble. Lors de la formation des apprentis-enseignants, nous préconisons qu'il soit explicité clairement aux futurs enseignants que :

- La préparation d'une photo ne dispense en aucun cas de la préparation d'une explication verbale. Il se peut que l'image ne s'affiche pas pour des raisons techniques, ou que l'apprenant ne comprenne pas le sens de l'item lexical en voyant la photo.
- L'image doit être commentée oralement, mais non par clavardage. Si l'apprenant doit regarder une photo et lire du texte en même temps, le processeur visuel du cerveau sera surchargé (Mayer, 2001).

A l'instar de Cicurel (2005) et sur le modèle de la flexibilité communicative, nous avons proposé le concept de **flexibilité multimodale** pour désigner la capacité à utiliser différents modes pour réaliser une même action, par exemple expliquer le sens d'un item lexical. Cette compétence inclut la capacité à jongler avec les modes pour renforcer son discours et/ou pour combler un mode défaillant. La flexibilité multimodale est utile pour des raisons techniques et pédagogiques. Si un canal de la plateforme cesse de fonctionner, l'enseignant est capable d'utiliser d'autres modes grâce à la compétence de flexibilité multimodale (Azaoui, 2017). Si un apprenant ne comprend pas une explication lexicale qui

est donnée dans un mode, par exemple par le biais d'images, l'enseignant devrait faire preuve d'une certaine flexibilité multimodale en donnant la même explication verbalement. Si une explication orale ne parvient pas à résoudre l'incompréhension, l'apprenti-enseignant devrait trouver un moyen d'utiliser ses expressions faciales ou ses gestes, ou bien de trouver un support multimédia sur Internet pour une expédition immédiate. Le développement de la flexibilité multimodale devrait être envisagé lors de la formation des enseignants. La compétence de la flexibilité multimodale n'est pas propre à l'enseignement par visioconférence : les enseignants doivent toujours savoir jongler entre parole, gestes, écriture et manipulation d'objets. Or la flexibilité multimodale est encore plus prégnante en visioconférence où il y a une palette étendue d'affordances, rendue évidente par le caractère multicanal de la plateforme.

10.2.3 Impliquer les apprenants

Nous avons constaté dans le chapitre 8 une forte implication des apprenants dans les séquences d'explication lexicale par les apprentis-enseignants. L'implication commence dès lors que l'apprenti-enseignant demande s'il y a des mots inconnus (cf. chapitre 7, exemple 12 en page 231). Ces ouvertures permettent aux apprenants de signaler une incompréhension immédiatement. Il revient cependant aux apprentis-enseignants de tenir compte des spécificités de la visioconférence, notamment le décalage entre l'émission et la réception des énoncés, afin que l'espace de collaboration ne soit pas clos de manière prématurée (cf. chapitre 8, exemple 2 en page 264). Hormis ces moments ponctuels, nous plaidons pour que les apprentis-enseignants signalent, dès la première séance, que les apprenants peuvent à tout moment demander une explication lexicale. Certains apprenants comme Naomi réussissent à utiliser le clavardage pour se faire aider par des apartés métalinguistiques, mais d'autres essaient sans succès de se faire aider, car ils ne rendent pas leur demande multimodalement perceptible et explicite. Nous avons aussi constaté l'importance qu'il y a à effectuer des rappels pour que l'apprenant montre qu'il a compris (cf. chapitre 7 exemple 4 en page 222). Ces constats rejoignent ceux de Fasel Lauzon (2014). Nous proposons que l'apprenti-enseignant n'accepte pas des réponses incertaines à des vérifications de compréhension. L'apprenant doit justifier qu'il a intégré de nouvelles connaissances. Les apprentis-enseignants devraient apprendre à repérer et à anticiper l'incompréhension chez l'apprenant en regardant les expressions faciales des apprenants et en tenant compte des hésitations (cf. chapitre 8 exemple 15 en page 287). L'intégration de ces compétences au répertoire

didactique (Cicurel, 2002) des enseignants de langues étrangères en visioconférence est à envisager.

Nous avons démontré qu'il existe plusieurs moyens multimodaux d'impliquer les apprenants, par exemple par des regards. Malgré l'impossibilité de maintenir le contact oculaire avec son interlocuteur par visioconférence, cette particularité a un avantage : un apprenti-enseignant peut donner l'impression à deux apprenants à la fois d'être regardés. L'apprenti-enseignant peut ainsi solliciter visuellement ses deux apprenants en même temps par un regard dirigé vers la webcam. Cependant, nous ignorons si l'effet stimulatoire dont parle De Chanay (2011) est ressenti par les apprenants de notre étude.

Puisque l'objectif du module de formation est de s'entraîner à l'oral, les noyaux polygérés des séquences d'explication lexicale sont à privilégier. Nous avons démontré que les apprentis-enseignants ouvrent des espaces de travail dans lesquels les pannes lexicales sont résolues collaborativement. Ces espaces sont ouverts par des rapprochements à l'écran, des commentaires et des gestes méta-sémiotiques, et des sourires. Nous avons élaboré la notion d'**alignement socioaffectif** pour désigner ces moments de partage d'émotions entre apprenti-enseignant et apprenant. Ces alignements socioaffectifs servent à réduire la gêne occasionnée par la correction. La notion de la charge d'implication, développée par Laufer & Hulstijn (2001) et reprise par Nation & Webb (2011) (voir annexe 3 en page 386), est limitée car elle ne prend pas en compte la dimension socioaffective de l'implication. La grille d'évaluation élaborée par ces auteurs mérite donc d'être améliorée par l'ajout de cette dimension et mise à l'épreuve d'études ultérieures.

10.2.4 Comparaisons interindividuelles

Notre corpus rend possible des comparaisons de plusieurs apprentis-enseignants dans des conditions similaires expliquant les mêmes items lexicaux. C'est une opportunité rare. Dans tous nos chapitres d'analyse, nous avons fait état d'une variété de pratiques multimodales mises en place par les apprentis-enseignants lors des séquences d'explication lexicale. En particulier, la comparaison du chapitre 9 entre plusieurs apprentis-enseignants en train d'expliquer un même item lexical a permis de dégager des différences qualitatives concernant l'utilisation des affordances de Visu. Tandis qu'Adèle et Mélissa semblent utiliser tous les canaux du dispositif, en particulier la webcam, d'autres comme Emilie, Etienne, Samia et Victor semblent préférer utiliser le microphone et le clavardage, la webcam étant une ressource moins sollicitée. Séverine n'utilise presque jamais la webcam pour communiquer.

Chaque apprenti-enseignant a donc sa propre grammaire individuelle et son propre profil multimodal, de même que chacun a son propre profil gestuel (Cadet & Tellier, 2007).

Nous disposons en outre de données quantitatives qui peuvent compléter ces ressentis. La quantité de données recueillies et annotées constitue un avantage majeur de notre corpus. Comme nous disposons de données pour sept apprentis-enseignants, y compris pour cinq au fil du semestre entier, nous pouvons tenter de mesurer d'éventuelles évolutions des pratiques d'enseignement. L'intérêt d'une telle analyse est double : d'une part, nous exploitons toutes les données dont nous disposons. D'autre part, nous pouvons tenter de répondre à l'une des questions que nous nous sommes posées dans le premier chapitre : un semestre est-il trop court pour mener une étude dite longitudinale ? Peut-on mesurer des évolutions des pratiques enseignantes en un semestre, ou faut-il plus de temps ?

Cohen (2017) a mené des analyses « longitudinales » similaires, en prenant le cadrage de la webcam comme unité d'analyse. Elle a défini le cadrage comme une échelle à cinq niveaux, allant d'un « très gros plan » où tout le champ de la webcam est rempli par le visage, jusqu'à un « plan américain » où le corps est éloigné de la webcam de sorte que la tête et les épaules soient pleinement visibles et ne remplissent pas la totalité du champ. Afin d'analyser le cadrage auprès de deux apprenties-enseignantes[1], Cohen a calculé le pourcentage du temps pour chaque plan pour quatre séances chacune. Elle a constaté une différence remarquable de l'utilisation de cette affordance entre les deux apprenties-enseignantes étudiées. Or, les pratiques de celles-ci n'ont pas changé de manière profonde du début à la fin du semestre.

Nous disposons des données quantitatives pour les sept critères suivants :

1. Le type de panne lexicale. La séquence d'explication lexicale survient-elle à la suite d'un problème de compréhension, de production, ou des deux à la fois ?
2. L'initiateur de la séquence d'explication lexicale. La séquence est-elle initiée par l'apprenti-enseignant, par l'apprenant, ou par les deux à la fois ?
3. La complexité des explications. Quelles sont les facettes de la connaissance de l'item lexical qui sont abordées ?

4-7. Quelles sont les ressources de Visu utilisées pour se focaliser sur la forme (4), sur le sens (5), sur l'usage (6), et pour impliquer l'apprenant (7) ?

Afin de ne pas surcharger le corps de texte de cette thèse, nous avons placé ces tableaux en annexe 6 en page 389.

[1] Adèle et Séverine

En regardant ces données, nous avons constaté une évolution de l'utilisation du clavardage chez Adèle et Victor pour se focaliser sur la forme des items lexicaux. Il semble que ces deux apprentis-enseignants deviennent progressivement plus à l'aise avec cet outil ou se rendent compte de son utilité pour répéter par écrit les items lexicaux qui font l'objet des explications. Or, cette évolution n'est qu'un ressenti : nos données ne nous permettent pas de réclamer une signification statistique. Pour cela, il aurait fallu des données contrôlées : par exemple, donner comme consigne aux apprentis-enseignants d'expliquer les mêmes mots dans les mêmes conditions. Comme chaque apprenant réagit différemment, il aurait fallu un expérimentateur qui fasse mine de ne pas comprendre les mots de la même façon. Nous sommes donc d'avis que les analyses qualitatives que nous avons menées au cours de cette thèse se prêtent mieux aux analyses didactiques. Il semble difficile de saisir les spécificités de l'enseignement des langues par visioconférence par le biais d'études expérimentales. Enfin, même si nous avions pu confirmer avec une signifiance statistique l'absence d'évolution, cela ne voudrait pas dire pour autant qu'un semestre soit trop court pour mener une étude longitudinale. Il se peut qu'un semestre soit de longueur suffisante pour une telle étude, mais que les apprentis-enseignants que nous avons observés n'aient pas évolué. Il se peut aussi que nous n'ayons pas défini des unités d'analyse susceptibles de faire émerger des évolutions.

Nos données obtenues pour les autres apprentis-enseignants et pour tous les autres critères corroborent les résultats de Cohen (2017) et ne montrent pas une évolution de pratiques multimodales. Nos résultats ne nous permettent pas de répondre à la question de la durée minimum nécessaire pour mener une étude longitudinale.

10.2 Limites de l'étude

Malgré la richesse de nos analyses, certaines limites liées à notre méthodologie et à nos analyses méritent d'être abordées.

10.2.1 Focalisation sur les apprenants

Hormis quelques exemples, nous n'avons pas analysé systématiquement et en détail les pratiques multimodales des apprenants, par exemple les moyens par lesquels ils signalent leur incompréhension des items lexicaux. En effet, nous avons parfois mené nos analyses comme si les canaux de la plateforme fonctionnaient dans un sens unique. Plusieurs chercheurs ont déjà mis au jour des moyens linguistiques par lesquels un interlocuteur peut signaler son incompréhension en faisant des demandes de reformulation, des hésitations

(Schegloff, Jefferson, & Sacks, 1977), ou des formulations provisoires (Lüdi, 1994). Nous avons montré quelques exemples où le visage de l'apprenant témoigne d'une certaine incompréhension (cf. chapitre 9 transcription 21 en page 336), et nous avons annoté les gestes produits par les apprenants lorsqu'ils décrivent l'item lexical recherché (cf. chapitre 7 exemple 25 en page 251). Mais une annotation plus complète permettrait de faire état de toutes ces stratégies et de les comptabiliser. L'élaboration d'une liste de pratiques multimodales des apprenants aiderait les enseignants en formation de mieux repérer les problèmes langagiers, notamment les pannes lexicales, que peuvent rencontrer les apprenants en visioconférence. Il se peut que les signaux émis par les apprenants en visioconférence ne soient pas facilement perceptibles par les enseignants peu expérimentés. Il se peut également que les apprenants ne soient pas conscients des signaux d'incompréhension qu'ils émettent. Une telle étude permettrait donc aussi à des professeurs de langue dont les apprenants suivent des formations complémentaires à distance de donner des conseils pour mieux signaler l'incompréhension et se faire aider.

Une annotation minutieuse de toutes les pratiques multimodales des apprenants permettrait de tester l'hypothèse de Sindoni (2013) selon laquelle la réponse à une question se fait par le même mode que celui par lequel elle est posée. Par exemple, si une demande d'explication se fait par clavardage, l'explication sera donnée par clavardage. Nos analyses montrent que ce n'est pas toujours le cas. Par exemple, Séverine répond oralement aux questions que son apprenante pose par clavardage, et Mélissa répond oralement et gestuellement à la demande d'aide d'Ana (cf. chapitre 7 exemple 25 en page 251).

Concernant l'utilisation de la webcam, l'utilisation d'un capteur qui suit et enregistre les mouvements oculaires permettrait de connaître exactement ce sur quoi les participants fixent leurs regards. Cette technologie coûte de moins en moins cher et devient accessible à tous comme les webcams. En effet, l'annotation des gestes des apprentis-enseignants n'est qu'une partie du travail. On sait déjà que les tous apprenants ne perçoivent pas les gestes pédagogiques de la même façon (Cadet & Tellier, 2007), et pour mieux comprendre leur efficacité, il serait bénéfique de savoir si tant est que les apprenants les regardent. L'utilisation de cette technologie pourrait enrichir la récolte des données lors d'études ultérieures. Si les apprentis-enseignants pouvaient savoir avec certitude quels gestes sont regardés par les apprenants, le travail réflexif des futurs enseignants sur leur propre pratique, ainsi que la conscientisation des gestes, serait facilité. Cela aiderait les apprentis-enseignant à travailler leur pratique gestuelle qui se développe non seulement avec l'évolution des différentes

méthodes d'enseignement (Cadet, 2014) mais aussi avec l'évolution des technologies. Le salon de rétrospection est un outil de grande valeur aux apprentis-enseignants car il leur permet de porter un regard critique sur leur propre pratique. La capacité à suivre les mouvements oculaires des apprenants enrichirait ce travail réflexif.

Enfin, chaque apprenant a une préférence modale pour l'apprentissage d'items lexicaux. Plass, Chun, Mayer, & Leutner (1998) ont analysé l'acquisition lexicale de sujets qui lisaient une histoire en langue étrangère sur un écran d'ordinateur et qui recevaient sur leur demande des explications verbales et/ou visuelles. Ils ont démontré expérimentalement que l'apprentissage est favorisé quand l'apprenant peut choisir son mode d'apprentissage, et quand le mode d'explication correspond au profil multimodal de l'apprenant. Connaître les profils des apprenants permettrait aux apprentis-enseignants de donner des explications sur mesure. Cela nécessiterait le développement de la compétence de la flexibilité multimodale. Il serait intéressant de voir si le résultat obtenu par Plass et al. (1998) concernant les préférences modales et les profils multimodaux est valable aussi bien pour les explications lexicales en visioconférence que pour les explications préenregistrées sur un écran.

10.2.2 Typologie des tâches

Nous n'avons pas élaboré une typologie des tâches pour savoir quels types de tâches provoquent le plus de travail collaboratif sur le lexique. Un tel développement permettrait de mesurer la rentabilité des tâches en termes de quantité de travail sur la langue, notamment sur le lexique, qui en résulte. Cependant, il semble que la majorité des tâches que nous avons observées soient propices à l'acquisition de connaissances lexicales comme en témoigne le nombre d'items lexicaux qui sont expliqués par plusieurs apprentis-enseignants, comme *bobo* et *goûter d'anniversaire*. En outre, nous n'avons pas comparé les séquences « planifiées » aux séquences « non planifiées ». Une analyse fine des plans de séances permettrait de classer les explications ainsi.

Nous n'avons pas mesuré l'apport de chaque explication lexicale au déroulement de la tâche ou à l'interaction. Guichon & Drissi (2008) définissent la pertinence d'une explication comme la valeur qu'elle ajoute à l'interaction. Comme le rappelle Fasel Lauzon (2014), une manière de mesurer la réussite d'une séquence d'explication lexicale consiste à voir si la tâche dont la réalisation nécessite la compréhension de l'item lexical a pu être complétée avec succès.

10.2.3 Efficacité des séquences d'explication lexicale

Cependant, il reste très difficile de mesurer l'efficacité des séquences d'explication lexicale. Fasel Lauzon (2014 : 206) souligne que si l'incompréhension est anticipée par l'enseignant, ce qui est le cas pour la moitié des séquences d'explication lexicale de notre corpus, « il est possible que le problème thématisé n'ait de caractère effectif pour aucun des participants ». Puis, même si l'apprenant dit avoir compris, ce n'est pas toujours le cas car l'apprenant peut feindre la compréhension « par volonté de ne pas prolonger la séquence ou par gêne de devoir admettre n'avoir toujours pas compris ». Il a d'ailleurs été démontré expérimentalement par Smith (2005) qu'*uptake* ne garantit pas acquisition. Nous ne nous sommes pas doté de moyens pour mesurer l'acquisition lexicale par les apprenants, par exemple un post-test. Une telle analyse permettrait de définir des critères d'analyse afin de mesurer l'efficacité de chaque séquence d'explication lexicale pour ensuite lister les facteurs qui semblent favoriser l'acquisition lexicale.

Afin de mesurer avec précision l'acquisition lexicale, il faudrait passer à un autre paradigme de recherche et administrer des post-tests aux participants. On pourrait demander aux étudiants à la fin de chaque séance de noter tous les items lexicaux qu'ils ont appris. Puis, tester la connaissance réceptive et la connaissance productive de plusieurs items lexicaux permettrait de mesurer la mémorisation de ceux-ci. Les résultats de post-tests, en combinaison avec nos annotations multimodales, permettraient de dire quelles méthodes d'explication et quels modes donnent les meilleurs résultats en termes d'apprentissage à l'instar d'Al-Seghayer (2001).[2] On pourrait aussi voir si l'intelligence spatiale des étudiants est corrélée à l'efficacité des explications multimodales, comme l'ont fait Fletcher & Tobias (2005). De telles analyses seraient possibles dans une démarche expérimentale.

10.2.4 Méthodologie et annotation multimodale

Nous avons commencé nos analyses en divisant les 295 séquences en trois phases selon le modèle canonique adopté (De Gaulmyn, 1991 ; Fasel Lauzon, 2008, 2014 ; Gülich, 1990a ; Lüdi, 1991). Ce modèle est influencé par l'analyse conversationnelle où les séquences d'explication lexicale sont considérées comme des réparations (Schegloff et al., 1977). Puis, nous avons fait le choix de nous focaliser sur les différentes facettes de la connaissance d'un

[2] Cet auteur a démontré expérimentalement que les explications multimodales sont plus efficaces que les explications seulement verbales (cf. chapitre 3 section 3.3.3.5 en page 99).

item lexical, notamment la forme et le sens. Si nous avions choisi d'autres critères d'analyse, par exemple la durée de chaque séquence d'explication, ou d'autres critères pris en compte dans le chapitre précédent, nous aurions obtenu des résultats différents. Cette décision méthodologique, prise en partie à cause du nombre élevé de séquences, nous permet de dire dans combien de séquences d'explication lexicale il y a focalisation sur quelle facette de l'item lexical, et par quels canaux de Visu. En raison de la nature écologique de notre corpus, nous ne sommes pas en mesure de dire globalement quelles stratégies ou quels canaux sont les plus efficaces. Un corpus contrôlé semble être nécessaire pour répondre à cette question (Tellier, 2013).

Notre façon d'annoter les données est strictement liée à nos questions de recherche. Nous avons décidé d'annoter la parole orale et écrite, les gestes, les expressions faciales, les regards, et les images envoyées. Ces critères d'analyse nous ont aidé à répondre à nos questions de recherche. Cependant, nous avons annoté uniquement les séquences d'explication lexicale que nous avons repérées. L'annotation de l'intégralité du corpus permettrait de comparer les pratiques multimodales des apprentis-enseignants lors des séquences d'explication lexicale aux pratiques de ceux-ci lors d'autres types de séquences. Une telle analyse serait extrêmement chronophage mais permettrait d'analyser les spécificités des séquences d'explication lexicale, et de voir si les profils multimodaux changent selon la situation, à l'instar de Tellier, Michel, & Wolff (2014).

La transcription des données n'est pas sans compromis car cette activité « effectue une transformation radicale des données temporelles et dynamiques en données écrites, spatialisées par conséquent » (Mondada, 2008 : 86). Nous avons commencé à transcrire avant de finaliser nos questions de recherche, ce qui constitue l'une des raisons pour laquelle nous avons choisi la convention ICOR au lieu de créer nous-même une convention de transcription. En effet, la convention de transcription oriente les résultats car la convention est créée en fonction des phénomènes que l'on défini *a priori* (Mondada, 2000). Plutôt que de modifier la convention ICOR, nous avons fait le choix d'ajouter une troisième colonne qui rend compte des phénomènes que nous avons repérés postérieurement à la transcription. Cette façon de représenter l'interaction facilite la lecture et guide l'attention sur un ou deux phénomènes à la fois. Or, cette représentation a ses limites. Il est impossible de lire la transcription dans la colonne à gauche et de lire en même temps la description textuelle sur la colonne de droite. Une convention de transcription plus élaborée aurait permis de le faire, mais cela aurait alourdi la transcription et nécessité un travail théorique avant de commencer la transcription.

La nature collaborative du projet ISMAEL, ainsi que la démarche inductive que nous avons adoptée, n'ont pas permis une telle entreprise.

Conclusion

Nos analyses nous ont permis de répondre à nos questions de recherche et de développer certaines notions. C'est en nous appuyant sur celles-ci que nous avons formulé certaines propositions didactiques pour améliorer la formation des enseignants. Cependant, nos analyses sont limitées dans certaines mesures. Ces limites ouvrent des pistes de recherche qui seront revisitées dans la conclusion.

Conclusion

Retour sur la démarche entreprise

Cette thèse s'est donné pour vocation d'étudier la multimodalité dans les échanges pédagogiques en visioconférence entre enseignants en formation et apprenants de niveau intermédiaire. Nous avons choisi comme entrée didactique le lexique en raison de l'intérêt personnel que nous y portons, à cause de son importance dans l'apprentissage langagier et parce que la plupart des déclencheurs de séquences de négociation du sens sont des items lexicaux.

Dans la première partie, nous avons dressé un état de l'art afin de nous doter d'outils théoriques nécessaires à l'analyse multimodale des séquences d'explication lexicale en visioconférence. La disponibilité à grande échelle d'outils de télécommunication en particulier l'Internet a changé la façon dont les humains communiquent, y compris la façon dont les langues étrangères sont enseignées. Ces évolutions ont donné naissance à un champ de recherche qui vise à étudier l'enseignement des langues étrangères à distance, notamment par visioconférence. Nous nous sommes inscrit dans cette lignée de recherche.

Nous avons justifié l'importance d'un enseignement systématique et explicite du vocabulaire en langue étrangère pour les niveaux débutant et intermédiaire. En effet, le lexique ne s'acquiert pas de manière naturelle en langue étrangère comme c'est le cas en langue maternelle. Il revient donc aux enseignants d'enseigner de manière systématique les premiers milliers de familles de mots, classées par ordre de fréquence, jusqu'à ce que les apprenants aient des compétences lexicales suffisantes pour être autonomes dans leur apprentissage. En nous appuyant sur des études expérimentales que nous avons résumées en page 72, nous avons dégagé certaines pratiques et conditions susceptibles de faciliter l'acquisition du lexique. L'élaboration de cette liste de pratiques a été nécessaire car notre corpus est écologique et aucun post-test n'a été administré aux participants. L'objectif de notre étude n'était donc pas de trouver les pratiques propices à l'acquisition du vocabulaire, mais de porter un regard critique sur les explications lexicales du corpus à partir des connaissances que nous avons développées en construisant l'état de l'art.

Puis, nous avons inscrit l'étude des séquences d'explication lexicale dans un courant de recherche plus large qui vise à étudier les séquences de négociation du sens qui

surviennent à la suite de pannes communicationnelles. Ce champ de recherche a été créé notamment par Schegloff, Jefferson et Sacks dans les années 1970. Ces études ont été appliquées par la suite au domaine de l'acquisition des langues par des chercheurs comme Long. La négociation du sens en langue étrangère est devenue objet d'études, et à partir de celles-ci, plusieurs modèles ont été élaborés pour rendre compte des phénomènes présents lors de chaque étape d'une séquence de négociation du sens. Nous avons pris comme exemple le modèle de Varonis & Gass (1985) en page 79 qui a été repris et élaboré pour la communication didactique par clavardage par Smith (2003). Nous avons rappelé que les séquences d'explication lexicale constituent un type de séquences de négociation du sens, et nous avons choisi un modèle canonique à trois phases (ouverture, noyau et clôture) (De Gaulmyn, 1991 ; Fasel Lauzon, 2008, 2014 ; Gülich, 1990a ; Lüdi, 1991). Nous avons critiqué ces études en mettant en évidence un manque de prise en compte de la dimension multimodale de la parole qui est pourtant fondamentale dans l'interaction humaine et a fortiori dans les échanges didactiques (Allen, 1995 ; Cadet & Tellier, 2007 ; Chanier & Vetter, 2006 ; Lazaraton, 2004 ; Tellier, 2014a).

Si on ne peut reprocher aux pionniers de l'analyse de la conversation, qui travaillaient sur des enregistrements audio en présentiel ou téléphonique, de ne pas avoir pris en compte la dimension multimodale de la parole, celle-ci devient tout particulièrement incontournable lorsque les interactions se passent sur une plateforme de communication multicanale comme la visioconférence. Nous avons donc argumenté en faveur d'une approche multimodale pour l'étude des séquences d'explication lexicale en visioconférence. Selon l'approche multimodale, chaque mode sémiotique contribue à la production du sens. Les interactants sélectionnent le mode ou les modes les plus aptes à faire passer le message. La hiérarchie des modes n'est pas stable, mais en constante réorganisation. Un mode devient plus ou moins dense selon son importance à un moment donné. A partir de ces bases théoriques, nous avons défini certains modes qui sont perceptiblement saillants et qui sont susceptibles d'être exploités par les apprentis-enseignants lors des séquences d'explication lexicale. Ceux-ci sont la parole orale et écrite, les gestes, les expressions faciales, les regards, les postures, et les images et ils sont véhiculés par les canaux de la plateforme Visu (le microphone, la webcam, le clavardage, et l'envoi de documents).

Dans la deuxième partie, nous avons détaillé la construction du corpus qui nous a permis d'étudier l'enseignement du FLE par visioconférence, notamment les explications lexicales. A partir de données écologiques qui ont été enregistrées en 2013, nous avons

collaboré à la construction d'un corpus audiovisuel de quinze heures et demie et de vingt-huit séances. Notre contribution majeure est la reconstruction de vingt-huit vidéos et la gestion de la transcription de celles-ci. Ce travail nous a permis de développer plusieurs compétences techniques, notamment la capacité de synchronisation de flux audiovisuels et de construction de montages vidéo. En respectant les enjeux éthiques de la recherche en sciences humaines, nous avons anonymisé le corpus en utilisant des pseudonymes et en floutant les images de certains apprenants. Puis, après avoir reconstruit les vingt-huit vidéos, nous les avons transcrites intégralement en utilisant le logiciel ELAN. La transcription nous a permis d'acquérir de l'expérience en tant que formateur, notamment auprès d'étudiants en Master et de chercheurs de divers pays lors d'un workshop à la conférence EUROCALL en 2016. [1] Le corpus et ses transcriptions ont pu être utilisés dans le projet ISMAEL (Guichon & Tellier, 2017).

A la suite de la construction et de la transcription des données, nous avons réfléchi au rapport que nous entretenons avec celles-ci en pensant que notre rapport à la langue, au sujet d'étude et aux participants nous offre un point de vue privilégié qui nous permet d'analyser les échanges. Puis, nous avons visionné la totalité des vingt-huit séances plusieurs fois afin de repérer toutes les séquences d'explication lexicale. Le choix d'entrer par le lexique s'est avéré être un bon choix car nous avons repéré 295 séquences d'explication lexicale. Nous les avons divisées en phases suivant le modèle canonique évoqué ci-dessus, et nous avons annoté les phénomènes multimodaux définis dans le troisième chapitre en utilisant le logiciel ELAN. Nous avons détaillé à l'aide d'exemples notre façon de représenter les annotations sur papier avec des tableaux à trois colonnes.

Dans la troisième partie, nous nous sommes efforcé d'analyser, sous plusieurs angles, la façon dont les apprentis-enseignants expliquent les items lexicaux par visioconférence. Nous avons dans un premier temps analysé la façon dont les canaux de Visu sont utilisés pour se focaliser sur la forme et sur le sens des items lexicaux. Nous avons repéré différents moyens linguistiques et multimodaux par lesquels les apprentis-enseignants se focalisent sur la forme afin de vérifier la compréhension, résoudre l'incompréhension, relier la forme au sens, corriger la prononciation, corriger l'orthographe, augmenter les compétences communicationnelles des apprenants, distinguer des mots phonétiquement proches, et donner des mots inconnus. Puis, nous avons analysé la façon dont les apprentis-enseignants se

[1] Lors de la conférence EUROCALL, qui s'est tenue à Limassol, à Chypre en 2016, nous avons mené un workshop avec Cathy Cohen, Eugénie Duthoit, Nicolas Guichon, Müge Satar et Ciara Wigham. Lors de ce workshop, nous avons initié les participants à l'utilisation d'ELAN.

focalisent sur le sens des items lexicaux en employant des explications en français, en recourant à la langue maternelle des apprenants, et en donnant des explications ostensives en utilisant des sons, des gestes, et des images.

Dans le chapitre 8, nous avons analysé la façon dont les apprentis-enseignants impliquent leurs apprenants dans les séquences d'explication lexicale. Les apprentis-enseignants impliquent leurs apprenants cognitivement en effectuant des rappels productifs et réceptifs, en personnalisant les définitions, et en ouvrant des espaces de travail collaboratif sur le lexique. Le côté socioaffectif est assuré par des sourires, rires, et partages d'émotions. Toutes les affordances de Visu, sauf l'envoi d'images, sont utilisées pour impliquer les apprenants.

Dans le chapitre 9, nous avons comparé quatre apprentis-enseignants qui expliquent « bobo » et cinq qui expliquent « goûter d'anniversaire ». Ce chapitre nous a permis de dégager des différences interindividuelles et de montrer que cinq apprentis-enseignants qui suivent le même plan de séance et qui se trouvent dans des conditions similaires expliquent différemment et ont des profils multimodaux variés. Ces observations qualitatives ont été complétées par des analyses quantitatives qui se trouvent en annexe 6 en page 389. Ces analyses ont montré d'une part que chaque apprenti-enseignant a son propre profil multimodal et d'autre part que ces profils ne semblent pas évoluer au fil du semestre. Cependant, l'usage du clavardage par Adèle et par Victor pour se focaliser sur la forme des items lexicaux semble être une pratique qui évolue du début à la fin du semestre.

Perspectives

Nous espérons que notre thèse a contribué à la didactique du FLE, notamment à la recherche sur les interactions didactiques en visioconférence. Afin de rendre compte d'une réalité particulière et éphémère, nous avons réemployé un modèle d'analyse de l'explication lexicale à un corpus d'interactions en visioconférence en adoptant une approche multimodale. L'approche adoptée nous a permis de mieux comprendre les enjeux de l'analyse multimodale des séquences d'explication lexicale. Nous avons démarré cette thèse en croyant que l'analyse de séquences d'explication lexicale consistait principalement à les diviser en phases et à inventorier les stratégies déployées par les acteurs lors de chaque phase. Or, l'objectif de l'analyse multimodale étant de démontrer qu'une combinaison de modes accroît la possibilité d'expression par rapport à un seul mode (Bateman, 2011), il devient indispensable d'associer les analyses à des objectifs communicatifs et/ou pédagogiques. C'est pourquoi nous avons

articulé nos analyses autour de nos questions de recherche concernant d'une part la transmission de la connaissance des items lexicaux et d'autre part l'implication des apprenants dans les séquences d'explication. Ces phénomènes nous semblent capitaux d'un point de vue didactique.

Dans le cadre de notre thèse, nous avons participé à la construction et à la transcription d'un grand corpus audiovisuel. Ce corpus a pu être exploité par une équipe de chercheurs et nous espérons que dans le futur, d'autres chercheurs en fassent autant. Le schéma d'annotation que nous avons élaboré est utilisable pour la communauté de chercheurs pour travailler sur d'autres corpus similaires. Nous avons actualisé l'annotation des gestes en visioconférence et avons représenté nos annotations en format papier en adoptant un schéma à trois colonnes. Nous espérons que notre façon d'annoter servira d'inspiration à d'autres chercheurs.

Il serait intéressant de mettre notre schéma d'annotation à l'épreuve d'une étude semi-contrôlée afin de mesurer des interactions possibles entre les phénomènes multimodaux annotés et l'acquisition lexicale. Pour cela, il faudrait créer une liste d'items lexicaux à enseigner et les intégrer dans une tâche. Il faudrait également contrôler les connaissances lexicales des apprenants en amont de la formation afin de vérifier que les items cibles sont inconnus. A la suite des interactions, un post-test permettrait de mesurer l'acquisition des items cibles par les apprenants. Il y a de fortes chances que la variabilité interindividuelle dont nous avons fait état à travers nos analyses se reproduira dans une étude semi-contrôlée. Cette variabilité permettrait de mesurer l'efficacité de certaines pratiques multimodales. Des analyses statistiques permettraient de dire que certaines pratiques se prêtent mieux que d'autres à l'acquisition lexicale en visioconférence. De telles expériences ont été menées par Al-Seghayer (2001), Kelly, McDevitt, & Esch, (2009), et par Kost, Foss, & Lenzini (1999) entre autres dans des contextes autres que celui d'enseignement par visioconférence. Une étude semi-contrôlée en visioconférence permettrait de mieux comprendre les spécificités de ce contexte et d'améliorer davantage la formation des enseignants.

Il serait possible aussi de mener une étude quantitative plus élaborée avec les données dont nous disposons actuellement. En effet, nous avons comptabilisé le nombre de séquences d'explication lexicale dans lesquelles il y a chaque phénomène multimodal que nous avons défini. Une étude supplémentaire nous permettrait de comptabiliser, à l'instar d'Inceoglu (2015), le nombre de gestes qu'il y a à l'intérieur de chaque séquence d'explication, et de comparer la production gestuelle des enseignants lors des explications déclenchées par

l'enseignant à la production gestuelle lors des séquences initiées par l'apprenant. Nous pourrions aussi catégoriser les items lexicaux (nom, verbe, adjectif, expression idiomatique, etc.) afin de faire émerger d'éventuelles interactions entre le type d'item lexical et les pratiques multimodales déployées par les apprentis-enseignants pour les expliquer.

Les notions que nous avons développées à travers nos analyses et résumées dans le chapitre précédent (les pléonasmes multimodaux, la flexibilité multimodale, et les alignements socioaffectifs) constituent d'autres pistes de recherche. Dans une démarche expérimentale, on pourrait mesurer l'impact de pléonasmes multimodaux sur l'apprentissage du vocabulaire. Selon le principe de redondance (Sweller, 2005b), la répétition des mêmes informations par deux modes nuit à l'apprentissage. Or, ce qui est redondant pour un expert ne l'est pas pour un novice. On pourrait tester notre hypothèse selon laquelle une focalisation sur les formes orale et écrite d'un item lexical n'aura pas les mêmes effets sur les locuteurs natifs que sur les non natifs. On peut formuler l'hypothèse que la répétition de la forme de l'item lexical par clavardage constituera une opération mentale *accessoire* pour un locuteur natif, et donc n'aidera pas à la compréhension du message, mais constituerait une opération mentale *essentielle* (Mayer & Moreno, 2003) pour un apprenant. Une telle étude permettrait de mettre une théorie cognitive de l'apprentissage multimédia à l'épreuve d'un corpus semi-contrôlée construit à partir d'interactions pédagogiques par visioconférence.

Nous avons montré dans le premier chapitre que les technologies de communication évoluent, et que l'enseignement des langues se transforme en conséquence. Il faut veiller cependant à ne pas refaire les mêmes analyses à chaque fois qu'un nouveau dispositif technique voit le jour. Demaizière (2000 : 321) remarqua il y a dix-huit ans que « Les microcosmes EAF[2], NTF[3], multimédia ont une propension regrettable à affadir certains concepts pour ensuite redécouvrir des évidences déjà mises en avant par des recherches antérieures (ou parallèles) ». L'explication des items lexicaux occupera toujours une place plus ou moins centrale dans l'enseignement des langues étrangères selon les courants méthodologiques, mais la visioconférence telle que nous la connaissons aujourd'hui n'est qu'une étape éphémère dans l'évolution constante de la communication médiée. Nous avons donc adopté une approche multimodale pour rendre compte des pratiques explicatives rendues possibles par la palette de ressources sémiotiques qui est propre à la visioconférence. Nous sommes d'avis que le développement de compétences sémio-pédagogiques nécessite une

[2] Nous supposons que ce sigle fait référence à l'Épreuve anticipée de français.
[3] Nouvelles technologies pour la formation

formation nourrie par la recherche ainsi qu'un travail réflexif de la part des apprentis-enseignants. Nous espérons que les analyses que nous avons présentées dans cette thèse, ainsi que les pistes de recherche que nous avons proposées à la fin, contribueront à la didactique du FLE et à la formation de futurs enseignants pour qui l'usage de technologies comme la visioconférence sera incontournable.

Bibliographie

Albero, B., & Thibault, F. (2009). La recherche française en sciences humaines et sociales sur les technologies en éducation. *Revue française de pédagogie. Recherches en éducation*, *169*, 53–66.

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of Visibility between Speaker and Listener on Gesture Production: Some Gestures Are Meant to Be Seen. *Journal of Memory and Language*, *44*, 169-188.

Allaf-Akbary, O., & Mosalli, Z. (2015). The effect of gestures on vocabulary acquisition and retention. *International Journal of Language Learning and Applied Linguistics World 5IJLLALW)*, *8*, 298-310.

Allen, L. Q. (1995). The effects of emblematic gestures on the development and access of mental representations of French expressions. *The Modern Language Journal*, *79*, 521–529.

Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: a comparative study. *Language Learning & Technology*, *5*, 202-232.

Altarriba, J., & Knickerbocker, H. (2011). Acquiring second language vocabulary through the use of images and words. In P. Trofimovich & K. McDonough (Éd.), *Applying Priming Methods to L2 Learning, Teaching and Research: 1* (p. 21-47). Amsterdam: John Benjamins.

Aston, G. (1986). Trouble-shooting in Interaction with Learners: the more the merrier? *Applied linguistics*, *7*, 128.

Azaoui, B. (2014). Multimodalité des signes et enjeux énonciatifs en classe de FL1/FLS. In *Le corps et la voix de l'enseignant : théorie et pratique* (p. 115-126). Paris: Ed. Maison des langues.

Azaoui, B. (2017). Faire face aux imprévus techniques. In N. Guichon & M. Tellier (Éd.), *Enseigner l'oral en ligne : Une approche multimodale* (p. 184-214). Paris: Les Éditions Didier.

Baddeley, A. (1993). *La mémoire humaine*. (S. Hollard, Trad.). Grenoble: Presses Universitaires de Grenoble.

Baldry, A., & Thibault, P. (2006). *Multimodal Transcription and Text Analysis*. London; Oakville, CT: Equinox Publishing.

Bange, P. (1992). À propos de la communication et de l'apprentissage de L2 (notamment dans ses formes institutionnelles). *Acquisition et interaction en langue étrangère*, *1*, 53–85.

Barcroft, J. (2006). Can writing a new word detract from learning it? More negative effects of forced output during vocabulary learning. *Second Language Research*, *22*, 487-497.

Bateman, J. (2011). The Decomposability of Semiotic Modes. In K. L. O'Halloran & B. A. Smith (Éd.), *Multimodal studies : exploring issues and domains* (p. 17-38). New York Routledge.

Beattie, G., & Shovelton, H. (1999). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica*, *123*, 1-30.

Benigno, V., Grossmann, F., & Kraif, O. (2015). Les collocations fondamentales: une piste pour l'apprentissage lexical. *Revue française de linguistique appliquée*, *20*, 81–96.

Bergmann, J. R. (1981). Ethnomethodologische Konversationsanalyse. In P. Schröder & H. Steger (Éd.), *Dialogforschung*. Düsseldorf, Schwann: Jahrbuch 1980 des Instituts für deutsche Sprache.

Bétrancourt, M., Guichon, N., & Prié, Y. (2011). Assessing the use of a Trace-Based Synchronous Tool for distant language tutoring (Vol. 1, p. 478-485). Présenté à 9th International Conference on Computer-Supported Collaborative Learning, Hong Kong.

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology*, *93*, 498-520.

Binon, J., & Verlinde, S. (2004). L'enseignement/apprentissage du vocabulaire et la lexicographie pédagogique du français sur objectifs spécifiques (FOS): le domaine du français des affaires. *Ela. Études de linguistique appliquée*, 271–283.

Blake, R. (2000). Computer mediated communication: A window on L2 Spanish interlanguage. *Language Learning & Technology*, *4*, 111-125.

Blin, F. (2016). The theory of affordances. In C. Caws & M.-J. Hamel (Éd.), *Language-Learner Computer Interactions: Theory, methodology and CALL applications* (p. 41-64). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Bogaards, P. (1994). *Le vocabulaire dans l'apprentissage des langues étrangères*. Saint-Jean de Braye: Les Éditions Didier.

Bonk, W. J. (2000). Second Language Lexical Knowledge and Listening Comprehension. *International Journal of Listening*, *14*, 14-31.

Bourdon-Larger, N. (2008). Démarche intégrative et parcours d'apprentissage modulable pour l'enseignement/apprentissage du lexique en classe de F.L.E. *Travaux de didactique du FLE*, *59*, 83-93.

Brett, A., Rothlein, L., & Hurley, M. (1996). Vocabulary acquisition from listening to stories and explanations of target words. *The Elementary School Journal*, *96*, 415–422.

Brouwer, C. E. (2003). Word Searches in NNS: NS Interaction: Opportunities for Language Learning? *Modern Language Journal*, 534–545.

Brudermann, C. (2010). Analyse de l'efficacité des stratégies de travail d'étudiants Lansad à distance dans un dispositif hybride – Étape d'une recherche-action. *Alsic*, *13*.

Cadet, L. (2014). Corps de l'enseignant : sujet de la recherche, objet de la formation. In *Le corps et la voix de l'enseignant : théorie et pratique* (p. 189-206). Paris: Ed. Maison des langues.

Cadet, L., & Tellier, M. (2007). Le geste pédagogique dans la formation des enseignants de langue étrangère. *Les Cahiers de Théodile*, *7*, 67-80.

Calaque, E. (2000). Enseignement et apprentissage du vocabulaire. *Lidil*, *21*, 17-36.

Cappellini, M., & Rivens Mompean, A. (2015). Role taking for teletandem pairs involved in multimodal online conversation: Some proposals for counseling practice. *Language Learning in Higher Education*, *5*, 243-264.

Cellule Corpus Complexes. Repéré à http://www.icar.cnrs.fr/ccc/

Celik, C. (2008). Analyse de pratiques de tutorat dans un campus numérique de maîtrise de français langue étrangère à distance. *Alsic*, *11*, 93-119.

Chanier, T., & Ciekanski, M. (2010). Utilité du partage des corpus pour l'analyse des interactions en ligne en situation d'apprentissage : un exemple d'approche méthodologique autour d'une base de corpus d'apprentissage. *Alsic*, *13*.

Chanier, T., & Vetter, A. (2006). Multimodalité et expression en langue étrangère dans une plate-forme audio-synchrone. *Alsic*, *9*.

Chanier, T., & Wigham, C. R. (2016). A scientific methodology for researching CALL interaction data: Multimodal LEarning and TEaching Corpora. In C. Caws & M.-J. Hamel (Éd.), *Language-Learner Computer Interactions: Theory, methodology and CALL applications* (p. 215-240). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Charlier, B., Deschryver, N., & Peraya, D. (2006). Apprendre en présence et à distance. *Distances et savoirs*, *4*, 469–496.

Chaudron, C. (1982). Vocabulary elaboration in teachers' speech to L2 learners. *Studies in Second Language Acquisition*, *4*, 170-180.

Chun, D. M., & Plass, J. L. (1997). Research on text comprehension in multimedia environments. *Language Learning & Technology*, *1*, 60-81.

Cicurel, F. (1993). Éléments d'un rituel communicatif dans les situations d'enseignement. In *Variations et rituels en classe de langue* (p. 22-54). Paris: Didier.

Cicurel, F. (2002). La classe de langue un lieu ordinaire, une interaction complexe. *Acquisition et interaction en langue étrangère*, *16*, 145–164.

Cicurel, F. (2005). La flexibilité communicative : un atout pour la construction de l'agit enseignant. *Le Français dans le monde. Recherches & applications*, 180-191.

Cicurel, F. (2011). *Les interactions dans l'enseignement des langues*. Paris: Les Éditions Didier.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*, 149-210.

Coady, J. (1993). Research on ESL/EFL Vocabulary Acquisition: Putting It In Context. In T. Huckin, M. Haynes, & J. Coady (Éd.), *Second language reading and vocabulary learning* (p. 3-23). New Jersey: Ablex Publishing Corporation.

Condreanu, T. (2014). *Analyse des comportements et expérience utilisateurs d'une plateforme de visioconférence pour l'enseignement des langues : Le cas de la phase 1.0 de VISU*. (Thèse de doctorat inédite). Université Lumière Lyon 2.

Codreanu, T., & Combe Celik, C. (2013). Effects of webcams on multimodal interactive learning. *ReCALL*, *25*, 30-47.

Codreanu, T., & Combe Celik, C. (2012). La médiation de l'interaction pédagogique sur une plateforme de visioconférence poste à poste. *Alsic*, *15*.

Codreanu, T., & Develotte, C. (2010). Enseigner par visioconférence poste à poste: cadre méthodologique pour l'analyse de pratiques tutorales. In *Actes des Journées Communication et Apprentissages Instrumentés en réseau (JOCAIR)* (p. 60-71). Amiens, France.

Cohen, C. (2017). Former à l'enseignement en ligne. In N. Guichon & M. Tellier (Éd.), *Enseigner l'oral en ligne* (p. 215-242). Paris: Didier.

Cohen, C., & Guichon, N. (2016). Analysing multimodal resources in pedagogical online exchanges. In C. Caws & M.-J. Hamel (Éd.), *Language-Learner Computer Interactions: Theory, methodology and CALL applications* (p. 187-213). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Colón de Carvajal, I. (2006). Modifications du cadre participatif dans des interactions médiatisées par un dispositif visiophonique. In *Innovations, usages, réseaux*. ATILF-CNRS. Repéré à https://edutice.archives-ouvertes.fr/edutice-00134909/

Coltier, D. (1988). Introduction et gestion des exemples dans les textes à thèse. *Pratiques*, *58*, 23-41.

Conseil de l'Europe. *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer*. Repéré à https://rm.coe.int/16802fc3a8

Cortier, C. (2006). De quelques enjeux et usages historiques du Français fondamental. *Documents pour l'histoire du français langue étrangère et seconde*, *36*, 1-3.

Cortier, C., & Faugeras, M.-P. (2004). Chapitre 5 : Développement des compétences morpho-lexicales en français langue seconde. In E. Calaque & J. David, *Didactique du lexique* (p. 73-87). Louvain-la-Neuve, Belgique: De Boeck Supérieur.

Cosnier, J., & Develotte, C. (2011). Le face à face en ligne, approche éthologique. In C. Develotte, R. Kern, & M.-N. Lamy (Éd.), *Décrire la conversation en ligne, Le face à face distanciel* (p. 27-50). Lyon: ENS Éditions.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671-684.

Craik, Fergus I. M., & Tulving, E. (1975). Depth of processing and the retntion of words in episodic memory. *Journal of Experimental Psychology*, *104*, 268-294.

Cuq, J.-P. (2003). *Dictionnaire de didactique du français : langue étrangère et seconde*. Paris: CLE Internationale.

Cuq, J.-P. (2004). Chapitre 4 : Le lexique en situation d'apprentissage guidé : pour une méthodologie d'enseignement interventionniste dans l'enseignement du français langue étrangère. In E. Calaque & J. David, *Didactique du lexique* (p. 61-72). Louvain-la-Neuve, Belgique: De Boeck Supérieur.

Dabène, L., Cicurel, F., Lauga-Hamid, M.-C., & Foerster, C. (1993). *Variations et rituels en classe de langue*. Paris: Didier.

De Chanay, H. (2011). La construction de l'éthos dans les conversations en ligne. In *Décrire la conversation en ligne, Le face à face distanciel* (p. 145-172). Lyon: ENS Éditions.

De Gaulmyn, M.-M. (1991). Expliquer des explications. In U. Dausendschön-Gay, E. Gülich, & U. Krafft (Éd.), *Linguistische Interaktionsanalysen* (p. 279-314). Tübingen: Max Niemeyer.

De Pietro, J.-F., Matthey, M., & Py, B. (1989). Acquisition et contrat didactique: les séquences potentiellement acquisitionnelles dans la conversation exolingue. In *Actes du troisième colloque régional de linguistique* (p. 99-119).

Degache, C., & Nissen, E. (2008). Formations hybrides et interactions en ligne du point de vue de l'enseignant : pratiques, représentations, évolutions. *Alsic*, *11*.

Dejean-Thircuir, C., Guichon, N., & Nicolaev, V. (2010). Compétences interactionnelles des tuteurs dans des échanges vidéographiques synchrones. *Distance et Savoirs*, *8*, 377-393.

Demaizière, F. (2000). Former les formateurs: nouveaux outils, nouvelles problématiques? In L. Duquette & M. Laurier (Éd.), *Apprendre une langue dans un environnement multimédia* (p. 305-323). Québec: Les Éditions LOGIQUES.

Demaizière, F. & Narcy-Combes, J-P. (2007). Du positionnement épistémologique aux données de terrain. Les cahiers de l'Acedle, 4, 1-20.

Develotte, C., & Drissi, S. (2013). Face à face distanciel et didactique des langues. *Le Français dans le monde. Recherches et applications*, 54–63.

Develotte, C., Guichon, N., & Kern, R. (2008). « Allo Berkeley ? Ici Lyon... Vous nous voyez bien ? » Étude d'un dispositif de formation en ligne synchrone franco-américain à travers les discours de ses usagers. *Alsic*, *11*, 129-156.

Develotte, C., Guichon, N., & Vincent, C. (2010). The use of webcam for teaching a foreign language in a desktop videoconferencing environment. *ReCALL*, *22*, 293-312.

Develotte, C., Kern, R., & Lamy, M.-N. (2011). *Décrire la conversation en ligne : Le face à face distanciel*. Lyon: ENS Editions.
Develotte, C., & Mangenot, F. (2007). Discontinuités didactiques et langagières au sein d'un dispositif pédagogique en ligne. *Glottopol*, 127-144.
Develotte, C., & Mangenot, F. (2010). Feed-back correctifs dans des formations de tuteurs de langue en ligne (en synchronie et en asynchronie). Présenté à Colloque international de l'AREF 2010, Genève.
Dobinson, T. (2001). Do learners learn from classroom interaction and does the teacher have a role to play? *Language teaching research*, *5*, 189–211.
Doughty, C., & Pica, T. (1986). "Information gap" tasks: Do they facilitate second language acquisition? *TESOL quarterly*, *20*, 305–325.
Drissi, S. (2011). *Apprendre à enseigner par visioconférence : étude d'interactions pédagogiques entre futurs enseignants et apprenants de FLE*. (Thèse de doctorat inédite). École Normale Supérieure de Lyon.
Ducancel, G. (1991). Expliquer à l'oral, à l'écrit, en sciences. *Repères*, *3*, 117-141.
Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, and coding. *Semiotica*, *1*, 49-98.
Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language learning*, *43*, 559–617.
Ellis, R. (1991). The Interaction Hypothesis: A Critical Evaluation. Présenté à Regional Language Centre Seminar, Singapore.
Ellis, R. (1995). Modified oral input and the acquisition of word meanings. *Applied Linguistics*, *16*, 409–441.
Ellis, R., Basturkmen, H., & Loewen, S. (2001). Learner uptake in communicative ESL lessons. *Language learning*, *51*, 281–318.
Ellis, R., Tanaka, Y., & Yamazaki, A. (1994). Classroom interaction, comprehension, and the acquisition of L2 word meanings. *Language learning*, *44*, 449–491.
Erten, İ. H., & Tekin, M. (2008). Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets. *System*, *36*, 407-422.
Fasel Lauzon, V. (2008). Interactions et apprentissages dans des séquences d'explication de vocabulaire. *Travaux neuchâtelois de linguistique*, *48*, 83-104.
Fasel Lauzon, V. (2014). *Comprendre et apprendre dans l'interaction : Les sequences d'explication en classe de francais langue seconde*. Bern: Peter Lang AG, Internationaler Verlag der Wissenschaften.
Fletcher, J. D., & Tobias, S. (2005). The Multimedia Principle. In R. E. Mayer (Éd.), *The Cambridge Handbook of Multimedia Learning* (p. 117-133). Cambridge: Cambridge University Press.
Flowerdew, J. (1992). Definitions in science lectures. *Applied linguistics*, *13*, 202–221.
Folse, K. S. (2004). *Vocabulary myths: applying second language research to classroom teaching*. Ann Arbor: The University of Michigan Press.
Gadoni, A., & Tellier, M. (2014). Conscientisation de la gestuelle pédagogique en formation initiale. In *Le corps et la voix de l'enseignant théorie et pratique* (p. 207-217). Paris: Ed. Maison des langues.
Galisson, R. (1983). Image et usage du dictionnaire chez des étudiants (en langue) de niveau avancé. *Études de Linguistique Appliquée*, *49*, 5-88.
Galisson, R. (1987). Accéder à la culture partagée par l'entremise des mots à C.C.P. *Études de Linguistique Appliquée*, *67*, 119-140.
Galisson, R. (1991). *De la langue à la culture par les mots*. Paris: CLE Internationale.

García, M. M. (2008). De l'importance des éléments suprasegmentaux dans l'apprentissage du lexique en L2. *Travaux de didactique du FLE*, *59*, 13-23.
Garfinkel, H. (1967). *Studies in ethnomethodology*. New Jersey: Prentice-Hall, Inc.
Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin Company.
Giroux, L. (2016). La place et le(s) rôle(s) de la langue maternelle des apprenants en cours de langue étrangère. *Synergies France*, *10*, 55-68.
Goodfellow, R., Jefferys, I., Milest, T., & Shirra, T. (1996). Face-to-face language learning at a distance? A study of a videoconference try-out. *ReCALL*, *8*, 5-16.
Goodwin, C. (1981). *Conversational organization: interaction between speakers and hearers*. New York: Academic Press.
Goodwin, M. H., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, *62*, 51-75.
Groot, P. (2000). Computer assisted second language vocabulary acquisition. *Language Learning & Technology*, *4*, 60-81.
Grossmann, F. (2011). Didactique du lexique: état des lieux et nouvelles orientations. *Pratiques. Linguistique, littérature, didactique*, *149-150*, 163–183.
Groupe ICOR (2013). *Convention ICOR*. Repéré à http://icar.univ-lyon2.fr/projets/corinte/documents/2013_Conv_ICOR_250313.pdf
Guichon, N. (2009). Training future language teachers to develop online tutotrs' competence through reflective analysis. *ReCALL*, *21*, 166-185.
Guichon, N. (2013). Une approche sémio-didactique de l'activité de l'enseignement de langue en ligne: réflexions méthodologiques. *Éducation & Didactique*, *7*, 101-116.
Guichon, N. (2017a). Se construire une présence pédagogique en ligne. In N. Guichon & M. Tellier (Éd.), *Enseigner l'oral en ligne : Une approche multimodale* (p. 29-58). Paris: Les Éditions Didier.
Guichon, N. (2017b). Sharing a multimodal corpus to study webcam-mediated language teaching. *Language Learning & Technology*, *21*, 55-74.
Guichon, N., Bétrancourt, M., & Prié, Y. (2012). Managing written and oral negative feedback in a synchronous online teaching situation. *Computer-Assisted Language Learning*, *25*, 181-197.
Guichon, N., & Cohen, C. (2014). The Impact of the Webcam on an Online L2 Interaction. *Canadian Modern Language Journal*, *70*, 331-354.
Guichon, N., & Cohen, C. (2016). Multimodality and CALL. In F. Farr & L. Murray (Éd.), *The Routledge Handbook of Language Learning and Technology.* (p. 509–521). Routledge.
Guichon, N., & Drissi, S. (2008). Tutorat de langue par visioconférence : comment former aux régulations pédagogiques ? *Les Cahiers de l'Acedle*, *5*, 1-32.
Guichon, N., & Nicolaev, V. (2009). Caractériser des tâches d'apprentissage et évaluer leur impact sur la production orale en L2. In *Actes du colloque EPAL 2009*. Grenoble.
Guichon, N., & Tellier, M. (Éd.). (2017). *Enseigner l'oral en ligne : une approche multimodale*. Paris: Les Éditions Didier.
Guichon, N., & Wigham, C. R. (2016). A semiotic perspective on webconferencing-supported language teaching. *ReCALL*, *28*, 62-82.
Gülich, E. (1990a). L'approche ethnométhodologique dans l'analyse du français parlé. In M. Charolles, S. Fisher, & J. Jayez (Éd.), *Le discours : représentations et interprétations* (p. 71-109). Nancy: Presses Universitaires de Nancy.

Gülich, E. (1990b). Pour une ethnométhodologie linguistique : description de séquences conversationnelles explicatives. In M. Charolles, S. Fisher, & J. Jayez (Éd.), *Le discours : représentations et interprétations* (p. 71-109). Nancy: Presses Universitaires de Nancy.

Gülich, E. (1991). Pour une ethnométhodologie linguistique: Description de séquences conversationnelles explicatives. In U. Dausendschön-Gay, E. Gülich, & U. Krafft (Éd.), *Linguistische Interaktionsanalysen* (p. 325-364). Tübingen: Max Niemeyer.

Gullberg, M. (2008). Gestures and second language acquisition. In P. Robinson & N. Ellis (Éd.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (p. 276-305). New York: Routledge.

Halliday, M. A. K.). (1978). *Language as social semiotic: the social interpretation of language and meaning*. London: E. Arnold.

Hatch, E. M. (1978). Discourse analysis and second language acquisition. In E. M. Hatch (Éd.), *Second language acquisition: A book of readings* (p. 401-435). Rowley, MA: Newbury House.

Hazenberg, S., & Hulstijn, J. H. (1996). Defining a minimal receptive second-language vocabulary for non-native university students: An empirical investigation. *Applied linguistics*, *17*, 145–163.

Heritage, J. (1988). Explanations as accounts: a conversation analytic perspective. In *Analyzing everyday explanation: a casebook of methods* (p. 127-144). London: Sage.

Hill, M., & Laufer, B. (2003). Type of task, time-on-task and electronic dictionaries in incidental vocabulary acquisition. *IRAL - International Review of Applied Linguistics in Language Teaching*, *41*.

Hilton, H. (2002). Modèles de l'acquisition lexicale en L2 : où en sommes-nous ? *Anglais de Spécialité (ASp)*, *35*, 201-217.

Hilton, H. (2003). L'accès au lexique mental dans une langue étrangère : le cas des francophones apprenant l'anglais. *Corela*, *2*, 25-43.

Holt, B. (2014). Stratégies multimodales mises en place par les apprentis-enseignants lors d'une interaction en visioconférence : le cas des séquences d'incompréhension. (Mémoire de Master 2 inédit). Université d'Aix-Marseille.

Holt, B., & Tellier, M. (2017). Conduire des explications lexicales. In N. Guichon & M. Tellier (Éd.), *Enseigner l'oral en ligne* (p. 59-90). Paris: Didier.

Holt, B., Tellier, M., & Guichon, N. (2015). The use of teaching gestures in an online multimodal environment: the case of incomprehension sequences. In *Gesture and Speech in Interaction 4th Edition* (p. 149-154). Nantes.

Horlacher, A.-S. (2015). *La dislocation à droite revisitée : une approche interactionniste*. Louvain-la-Neuve: De Boeck Duculot.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *Studies in second language acquisition*, *21*, 181–193.

Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The modern language journal*, *80*, 327–339.

Hummel, K. (2010). Translation and short-term vocabulary retention: Hinderance or help? *Language Teaching Research*, *14*, 61-74.

Inceoglu, S. (2015). Teacher Gesture and Lexical Focus on Form in a Foreign Language Classroom. *Canadian Modern Language Review*, *71*, 130-154.

Jefferson, G. (1972). Side Sequences. In D. Sudnow (Éd.), *Studies in Social Interaction* (p. 294-338). New York.

Jewitt, C. (2014a). An introduction to multimodality. In C. Jewitt (Éd.), *The Routledge Handbook of Multimodal Analysis* (Second Edition, p. 15-30). London: Routledge.

Jewitt, C. (2014b). Different approaches to multimodality. In C. Jewitt (Éd.), *The Routledge Handbook of Multimodal Analysis* (Second Edition, p. 31-43). London: Routledge.

Jewitt, C. (2014c). Multimodal approaches. In S. Norris & C. D. Maier (Éd.), *Interactions, images and texts: a reader in multimodality* (p. 127-136). Boston/Berlin: De Gruyter Mouton.

Jézégou, A. (2010). Créer de la présence à distance en e-learning. Cadre théorique, définition, et dimensions clés. *Distances et savoirs*, *8*, 257-274.

Joe, A. (1995). Text-based tasks and incidental vocabulary learning. *Second Language Research*, *11*, 149–158.

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied linguistics*, *19*, 357–377.

Jones, R. (2014). Mediated discourse analysis. In S. Norris & C. D. Maier (Éd.), *Interactions, images and texts: a reader in multimodality* (p. 39-51). Boston/Berlin: De Gruyter Mouton.

Jones, R. H. (2004). The Problem of Context in Computer-Mediated Communication. In P. Le Vine & R. Scollon (Éd.), *Discourse and technology: multimodal discourse analysis* (p. 20-33). Washington DC Georgetown University Press.

Kelly, N., & Bruen, J. (2015). Translation as a pedagogical tool in the foreign language classroom: A qualitative study of attitudes and behaviours. *Language Teaching Research*, *19*, 150-168.

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two Sides of the Same Coin: Speech and Gesture Mutually Interact to Enhance Comprehension. *Psychological Science*, *21*, 260-267.

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*, 313-334.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kerbrat-Orecchioni, C.), & Cosnier, J. (1987). *Décrire la conversation*. Lyon: Presses Universitaires de Lyon.

Kim, Y. (2008). The contribution of collaborative and individual tasks to the acquisition of L2 vocabulary. *The Modern Language Journal*, *92*, 114–130.

Kost, C. R., Foss, P., & Lenzini, J. J. (1999). Textual and Pictorial Glosses: Effectiveness on Incidental Vocabulary Growth When Reading in a Foreign Language. *Foreign Language Annals*, *32*, 89-97.

Kramsch, C. J. (1979). Word watching: Learning vocabulary becomes a hobby. *Foreign Language Annals*, *12*, 153-158.

Krashen, S. (1985). *The input hypothesis: Issues and implications*. New York: Longman.

Kress, G. (2014). What is mode? In C. Jewitt (Éd.), *The Routledge Handbook of Multimodal Analysis* (Second Edition, p. 60-75). London: Routledge.

Kress, G., & Van Leeuwen, T. (2001). *Multimodal discourse: the modes and media of contemporary communication*. London: Hodder Education.

Kupetz, M. (2011). Multimodal resources in students' explanations in CLIL interaction. *Novitas-Royal*, *5*, 121-142.

Labov, W. (1973). The linguistic consequences of being a lame. *Language in society*, *2*, 81–115.
Labov, W. (1976). *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphie: University of Pennsylvania Press.
Lamy, M.-N. (2006). Conversations multimodales : l'enseignement-apprentissage de l'oral à l'heure des écrans partagés. *Le français dans le monde*, *Recher(40)*, 129-138.
Lancien, T. (1998). Images mobiles et multimédia. *Études de linguistique appliquée : revue de didactologie des langues-cultures*, *Hypermédia et apprentissage des langues*, 171-182.
Lascar, J. & Valéro, D. (2016). *Présentation du logiciel Transana*. Communication présentée à l'Action National de Formation : Collecter et produire des données pour la recherche en SHS, Frejus. Repéré à https://anf-mate2016.sciencesconf.org/data/MATE_SHS_2016_LascarJ_ValeroD.pdf
Laufer, B. (1994). Appropriation du vocabulaire: mots faciles, mots difficiles, mots impossibles. *Acquisition et interaction en Langue Etrangère*, *3*, 97–113.
Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: the construct of task-induced involvement. *Applied Linguistics*, *21*, 1-26.
Laufer, B., & Shmueli, K. (1997). Memorizing new words: Does teaching have anything to do with it? *RELC journal*, *28*, 89–108.
Lazaraton, A. (2004). Gesture and speech in the vocabulary explanations of one ESL teacher: A microanalytic inquiry. *Language learning*, *54*, 79–117.
Leeman, D. (2005). Le vertige de l'infini ou de la difficulté de didactiser le lexique. *Le français aujourd'hui*, *148*, 89-99.
Long, M. (1981). Input, interaction, and second-language acquisition. *Native Language and Foreign Language Acquisition*, *379*, 259-278.
Long, M. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, *4*, 126-141.
Long, M. (1991). Focus on Form: A Design Feature in Language Teaching Methodology. In K. de Bot, R. B. Ginsberg, & C. Kramsch (Éd.), *Foreign Language Research in Cross-Cultural Perspective* (p. 39-52). Amsterdam/Philadelphia: John Benjamins Publishing Company.
Lotto, L., & De Groot, A. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language learning*, *48*, 31–69.
Low, R., & Sweller, J. (2005). The Modality Principle in Multimedia Learning. In R. E. Mayer (Éd.), *The Cambridge Handbook of Multimedia Learning* (p. 147-158). Cambridge: Cambridge University Press.
Lüdi, G. (1991). Construire ensemble les mots pour le dire. A propos de l'origine discursive des connaissances lexicales. In U. Dausendschön-Gay, E. Gülich, & U. Krafft (Éd.), *Linguistische Interaktionsanalysen* (p. 193-224). Tübingen: Max Niemeyer.
Lüdi, G. (1994). Dénomination médiate et bricolage lexical en situation exolingue. *Acquisition et interaction en langue étrangère*, *3*, 115–146.
Macedonia, M. (2014). Bringing back the body into the mind: gestures enhance word learning in foreign language. *Frontiers in Psychology*, *5*.
Macedonia, M., Müller, K., & Friederici, A. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, *32*, 982-998.

Mangenot, F. (2008). Pratiques pédagogiques instrumentées et propriétés des outils: le cas des forums. *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation (STICEF)*, *15*.

Mangenot, F. (2011). Spécificités du tutorat en langues. In C. Depover, B. De Lièvre, D. Peraya, J.-J. Quintin, & A. Jaillet (Éd.), *Le tutorat en formation à distance* (p. 213-225). Bruxelles: De Boeck Supérieur.

Mangenot, F., & Moulin, C. (2000). Apprentissage lexicaux et technologies de l'information et de la communication. *Lidil*, *21*, 121-139.

Marijanović, V., Billières, M., & Panissal, N. (2008). L'identification du mot écrit en FLE auprès des apprenants débutants. *Travaux de didactique du FLE*, *59*, 25-33.

Marra, A., & Pallotti, G. (2006). Les logonymes dans la classe de langue. In M. Faraco (Éd.), *La classe de langue. Théories, méthodes et pratiques* (p. 129-151). Aix-en-Provence: Presses de l'Université de Provence.

Mayer, R. E. (2001). *Multimedia Learning*. Cambridge: Cambridge University Press.

Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, *38*, 43-52.

McDonough, K., & Sunitham, W. (2009). Collaborative dialogue between Thai EFL learners during self-access computer activities. *TESOL quarterly*, *43*, 231–254.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, *92*, 350-371.

McNeill, D. (1992). *Hand and Mind*. Chicago: The University of Chicago Press.

McNeill, D. (2005). *Gesture & Thought*. Chicago: The University of Chicago Press.

Merlino, S., & Traverso, V. (2009). Les séquences de traduction spontanée comme mécanisme de réparation dans des interactions professionnelles. *Synergies Pays Germanophones, Gerflint*, 129-143.

Miecznikowski, J. (2005). *Le traitement de problèmes lexicaux lors de discussions scientifiques en situation plurilingue : Procédés interactionnels et effets sur le développement du savoir*. Berne: Peter Lang.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2011). Seeing and Being Seen: The Effects on Gesture Production. *Journal of Computer-Mediated Communication*, *17*, 77-100.

Mondada, L. (2000). Les effets théoriques des pratiques de transcription. *Linx*, *42*, 131-146.

Mondada, L. (2008). La transcription dans la perspective de la linguistique interactionnelle. In M. Bilger (Éd.), *Données orales : les enjeux de la transcription*. Perpignan: Presses Universitaires de Perpignan.

Mortureux, M.-F. (2008). *La lexicologie entre langue et discours* (2e édition revue et actualisée). Paris: Armand Colin.

Nation, I. S. P. (2013). *Learning Vocabulary in Another Language, Second Edition* (Cambridge University Press). Cambridge.

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston: Heinle.

Nation, P. (2000). Using communicative tasks to teach vocabulary. *Guidelines*, *22*, 15-19.

Newton, J. (2001). Options for vocabulary learning through communication tasks. *ELT Journal*, *55*, 30-37.

Newton, J. (1995). Task-based interaction and incidental vocabulary learning: A case study. *Second Language Research*, *11*, 159–176.

Newton, J. (2013). Incidental vocabulary learning in classroom communication tasks. *Language Teaching Research*, *17*, 164-187.

Nicolaev, V. (2010). Les négociations de sens dans un dispositif d'apprentissage des langues en ligne synchrone par visioconférence. *Les Cahiers de l'Acedle*, *7*, 169–198.

Nicolaev, V. (2012). *L'apprentissage du FLE dans un dispositif vidéographique synchrone: étude des séquences métalinguistiques*. (Thèse de doctorat inédite). École Normale Supérieure de Lyon.

Nissen, E. (2006). Scénarios de communication en ligne dans des formations hybrides. *Le Français dans le monde. Recherches et applications*, *« Les échanges en ligne dans l'apprentissage et la formation »*, *numéro spécial*, *juillet 2006*, 44–58.

Norris, S. (2004). *Analyzing Multimodal Interaction: A Methodological Framework*. New York: Routledge.

Norris, S. (2014). Modal density and modal configurations. In C. Jewitt (Éd.), *The Routledge Handbook of Multimodal Analysis* (Second Edition, p. 86-99). London: Routledge.

O'Dowd, R. (2000). Intercultural learning via videoconferencing: A pilot exchange project. *ReCALL*, *12*, 49–61.

O'Dowd, R. (2005). The Use of Videoconferencing and E-mail as Mediators of Intercultural Student Ethnography. In J. A. Belz & S. L. Thorne (Éd.), *Internet-mediated Intercultural Foreign Language Education* (p. 2-30). Boston: Thomson Higher Education.

O'Halloran, K. L., & Smith, B. A. (2011). *Multimodal studies : exploring issues and domains*. New York: Routledge.

Paivio, A. (2006). *Mind and its evolution; A dual coding theoretical interpretation*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Pekarek Doehler, S. (2001). Dislocation à gauche et organisation interactionnelle. *Marges linguistiques*, *2*, 177–194.

Pekarek Doehler, S. (2004). Une approche interactionniste de la grammaire : réflexions autour du codage grammatical de la référence et des topics chez l'apprenant avancé d'une L2. *AILE Acquisition et Interaction en Langue Etrangère*, *21*, 123-166.

Peterson, M. (2010). Task-Based Language Teaching in Network-Based CALL: An Analysis of Research on Learner Interactions in Synchronous CMC. In M. Thomas & H. Reinders (Éd.), *Task-Based Language Learning and Teaching with Technology* (p. 41-62). London: Continuum International Publishing Group.

Pica, T. (1994). Research on Negotiation: What Does It Reveal About Second-Language Learning Conditions, Processes, and Outcomes? *Language learning*, *44*, 493–527.

Pica, T., Young, R., & Doughty, C. (1987). The impact of interaction on comprehension. *Tesol Quarterly*, *21*, 737–758.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second language multimedia learning environment. *Journal of Educational Psychology*, *90*, 25-36.

Polguère, A. (2004). Chapitre 8 : La paraphrase comme outil pédagogique de modélisation des liens lexicaux. In E. Calaque & J. David, *Didactique du lexique* (p. 115-125). Louvain-la-Neuve, Belgique: De Boeck Supérieur.

Porquier, R. (1979). Stratégies de communication en langue non-maternelle. *Travaux du Centre de Recherches Sémiologiques*, *33*, 38-52.

Porquier, R. (1984). Communication exolingue et apprentissage des langues. In B. Py (Éd.), *Encrages : acquisition d'une langue étrangère* (p. 17-47). Université de Paris VIII et de Neuchâtel.

Porquier, R., & Py, B. (2008). *Apprentissage d'une langue étrangère : contextes et discours*. Le mesnil-sur-l'Estrée: Imprimerie Nouvelle Firmin Didot.

Pressley, M., Levin, J., & McDaniel, M. (1987). Remembering versus inferring what a word means: Mnemonic and contextual approaches. In M. McGeown & M. E. Curtis (Éd.), *The nature of vocabulary acquisition*. Hillsdale, N.J.: Erlbaum.

Rabardel, P. (1995). *Les hommes & les technologies*. Paris: Armand Colin.

Rey-Debove, J. (1997). *Le métalangage : étude linguistique du discours sur le langage*. Paris: Armand Colin.

Rivens Mompean, A., & Cappellini, M. (2015). Teletandem as a complex learning environment: Looking for a model. *DELTA: Documentação de Estudos Em Lingüística Teórica e Aplicada*, *31*, 633-663.

Rivens Mompean, A., & Guichon, N. (2009). Assessing the use of aids for a computer-mediated task: taking notes while listening. *JALT CALL Journal*, *5*, 45-60.

Rivière, V., & Guichon, N. (2014). Construction de bilans rétroactifs par des apprentis tuteurs de langue en ligne: régimes d'action et dynamiques sociocognitives. *Le Français dans le Monde Recherches et Applications*, 118–135.

Robert, J.-P. (2008). *Didctionnaire pratique de didactique du FLE*. Paris: Éditions Ophrys.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, *50*, 696-735.

Sarré, C. (2011). Computer-mediated negotiated interactions: how is meaning negotiated in discussion boards, text chat and videoconferencing? In S. Thouësny & L. Bradley (Éd.), *Second language teaching and learning with technology: views of emergent researchers* (p. 189-210). Dublin: Research-publishing.net.

Satar, M. (2013). Multimodal language learner interactions via desktop videoconferencing within a framework of social presence: Gaze. *ReCALL*, *25*, 122-142.

Satar, M. (2015). Sustaining multimodal language learner interactions online. *CALICO Journal*, *32*, 480-507.

Scheflen, A. E. (1964). The significance of posture in communication systems. *Psychiatry*, *27*, 316-331.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The Preference for Self-Correction in the Organization of Repair in Conversation. *Language*, *53*, 361-382.

Schmidt, R. W. (1990). The role of consciousness in second language learning1. *Applied linguistics*, *11*, 129–158.

Schmitt, N. (2008). Teaching vocabulary. *White Plains, NY: Pearson Education*.

Sime, D. (2006). What do learners make of teachers' gestures in the language classroom? *IRAL - International Review of Applied Linguistics in Language Teaching*, *44*, 211-230.

Sime, D. (2008). "Because of Her Gesture, It's Very Easy to Understand"—Learners' Perceptions of Teachers' Gestures in the Foreign Language Class. In S. McCafferty & G. Stam (Éd.), *Gesture: Second Language Acquisition and Classroom Research* (p. 259-279). New York: Routledge.

Sindoni, M. G. (2013). *Spoken and Written Discourse in Online Interactions: A Multimodal Approach*. New York, NY: Routledge.

Singleton, D. (1994). Introduction: le rôle de la forme et du sens dans le lexique mental en L2. *Acquisition et interaction en langue étrangere*, 3–27.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied linguistics*, *17*, 38–62.

Smith, B. (2003). Computer-Mediated Negotiated Interaction: An Expanded Model. *The Modern Language Journal*, *87*, 38-57.

Smith, B. (2005). The Relationship Between Negotiated Interaction, Learner Uptake, and Lexical Acquisition in Task-Based Computer-Mediated Communication. *Tesol Quarterly*, *39*, 33–58.

Soubrié, T. (2008). La difficile articulation du présentiel et de la distance dans le cadre d'un cours hybride en master. *Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication*, *11*.

Soubrié, T. (2016). La pratique réflexive, une dimension centrale dans la formation des enseignants au numérique. In C. Ollivier, T. Gaillat, & L. Puren (Éd.), *Numérique et formation des enseignants de langue : Pistes et imaginaires* (p. 1-18). Paris: Éditions des archives contemporaines.

Stahl, S. A., & Clark, C. H. (1987). The effects of participatory expectations in classroom discussion on the learning of science vocabulary. *American Educational Research Journal*, *24*, 541–555.

Stahl, S. A., & Vancil, S. J. (1986). Discussion is what makes semantic maps work in vocabulary instruction. *The Reading Teacher*, *40*, 62–67.

Stam, G. (2007). Second Language Acquisition from a McNeillian Perspective. In S. D. Duncan, J. Cassell, & E. T. Levy (Éd.), *Gesture and the dynamic dimension of language: Essays in honor of David McNeill* (p. 117-124). Amsterdam: John Benjamins Publishing Company.

Stam, G., & McCafferty, S. (2008). Gesture Studies and Second Language Acquisition: A Review. In S. McCafferty & G. Stam (Éd.), *Gesture: Second Language Acquisition and Classroom Research* (p. 3-24). New York: Routledge.

Stöckl, H. (2014). Semiotic paradigms and multimodality. In C. Jewitt (Éd.), *The Routledge Handbook of Multimodal Analysis* (Second Edition, p. 274-286). London: Routledge.

Stockwell, G. (2010). Effects of Multimodality in Computer-Mediated Communication Tasks. In M. Thomas & H. Reinders (Éd.), *Task-Based Language Learning and Teaching with Technology* (p. 83-104). London: Continuum International Publishing Group.

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, *55*, 661–699.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Éd.), *Input in second language acquisition* (p. 235-252). Rowley, MA: Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In *Principles and Practice in Applied Linguistics: Studies in Honour of H.G. Widdowson* (p. 125-144). Oxford: Oxford University Press.

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Éd.), *Sociocultural theory and L2 learning* (p. 97-114). Oxford: Oxford University Press.

Sweller, J. (2005a). Implications of Cognitive Load Theory for Multimedia Learning. In R. E. Mayer (Éd.), *The Cambridge Handbook of Multimedia Learning* (p. 19-30). Cambridge: Cambridge University Press.

Sweller, J. (2005b). The Redundancy Principle in Multimedia Learning. In R. E. Mayer (Éd.), *The Cambridge Handbook of Multimedia Learning* (p. 159-167). Cambridge: Cambridge University Press.

Tellier, M. (2005). L'utilisation des gestes en classe de langue : Comment évaluer leur effet sur la mémorisation du lexique ? In M. Billières, P. Gaillard, & N. Spanghero-Gaillard (Éd.). Présenté à Colloque International de Didactique Cognitive, Toulouse.

Tellier, M. (2008a). Dire avec des gestes. *Le français dans le monde: recherche et application*, 40-50.

Tellier, M. (2008b). The effect of gestures on second language memorisation by young children. *Gesture*, *8*, 219-235.

Tellier, M. (2009a). The development of gesture. In K. de Bot & R. W. Schrauf (Éd.), *Language Development Over the Lifespan* (p. 191-216). New York: Routledge.

Tellier, M. (2009b). Usage pédagogique et perception de la multimodalité pour l'accès au sens en langue étrangère. In R. Bergeron, G. Plessis-Bélair, & L. Lafontaine (Éd.), *La place des savoirs oraux dans le contexte scolaire d'aujourd'hui* (p. 223-245). Montréal: Presses Universitaires du Québec.

Tellier, M. (2012). Former à l'étude de la gestuelle : réflexions didactique. In R. Vion, A. Giacomi, & C. Vargas (Éd.), *La corporalité du langage : Multimodalité, discours et écriture* (p. 73-85). Aix-en-Provence: Press Universitaires de Provence.

Tellier, M. (2013). De l'usiage du corpus semi-contrôlé dans la recherche en didactique des langues. *Les Cahiers de l'ASDIFLE*, *FLE : L'instant et l'histoire*, *Actes des 49e et 50e rencontres*, 39-47.

Tellier, M. (2014a). Donner du corps à son cours. In *Le corps et la voix de l'enseignant : théorie et pratique* (p. 101-114). Paris: Ed. Maison des langues.

Tellier, M. (2014b). Quelques orientations méthodologiques pour étudier la gestuelle dans des corpus spontanés et semi-contrôlés. *Discours*, *15*.

Tellier, M., & Cadet, L. (2014). *Le corps et la voix de l'enseignant : théorie et pratique*. Paris: Editions Maison des langues.

Tellier, M., Michel, L., & Wolff, L. (2014). Variations inter et intra-individuelles de la gestuelle chez l'enseignante de maternelle. In *Le corps et la voix de l'enseignant : théorie et pratique* (p. 109-122). Paris: Éditions Maison des langues.

Tellier, M., & Stam, G. (2010). Découvrir le pouvoir de ses mains : La gestuelle des futurs enseignants de langue. In *Spécificités et diversité des interactions didactiques : disciplines, finalités, contextes* (p. 1-4). Lyon, France.

Tellier, M., & Stam, G. (2012). Stratégies verbales et gestuelles dans l'explication lexicale d'un verbe d'action. In V. Rivière (Éd.), *Spécificités et diversité des interactions didactiques* (p. 357-374). Paris: Riveneuve éditions.

Ten Have, P. (1999). *Doing Conversation Analysis: A Practical Guide*. London: SAGE Publications Ltd.

Théophanous, O. (2001). Déviations lexicales de forme et de sens chez les apprenants de français langue étrangère. *Revue Française de Linguistique Appliquée*, *6*, 107-120.

Théophanous, O. (2008). Polylexicalité et préconstruction lexicale : pour une vision élargie du lexique en FLE. *Travaux de didactique du FLE*, 45-58.

Thorne, S. L. (2005). Pedagogical and Praxiological Lessons from Internet-mediated Intercultural Foreign Language Education Research. In J. A. Belz & S. L. Thorne (Éd.), *Internet-mediated Intercultural Foreign Language Education* (p. 2-30). Boston: Thomson Higher Education.

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second language research*, *13*, 138–163.

Traverso, V. (2011). Chevauchements de parole, résolution et réparation dans la conversation en ligne. In *Décrire la conversation en ligne, Le face à face distanciel* (p. 117-143). Lyon: ENS Éditions.

Tréville, M.-C. (2000). *Vocabulaire et apprentissage d'une langue seconde : recherches et théories*. Québec: Les Éditions LOGIQUES.

Tréville, M.-C., & Duquette, L. (1996). *Enseigner le vocabulaire en classe de langue*. Paris: Hachette F.L.E.

Trimble, L. (1985). *English for Science and Technology: A Discourse Approach*. Cambridge: Cambridge University Press.

Tudini, V. (2003). Éléments conversationnels du clavardage : un entraînement à l´expression orale pour les apprenants de langues à distance ? *Alsic*, *6*, 63-81.

Van der Linden, E. (2006). Lexique mental et apprentissage des mots. *Revue Française de Linguistique Appliquée*, *11*, 33-44.

Van Leeuwen, T. (2004). Ten Reasons Why Linguists Should Pay Attention to Visual Communication. In P. Le Vine & R. Scollon (Éd.), *Discourse and technology: multimodal discourse analysis* (p. 7-19). Washington DC Georgetown University Press.

Van Leeuwen, T. (2005). *Introducing social semiotics*. London: Routledge.

Varonis, E. M., & Gass, S. (1985). Non-native/Non-native Conversations: A Model for Negotiation of Meaning. *Applied Linguistics*, *6*, 71-90.

Vidal, J., & Wigham, C. R. (2017). Fournir des rétroactions en ligne. In N. Guichon & M. Tellier (Éd.), *Enseigner l'oral en ligne : Une approche multimodale* (p. 125-150). Paris: Les Éditions Didier.

Vincent, C. (2012). *Interactions pédagogiques "fortement multimodales" en ligne : le cas de tuteurs en formation*. (Thèse de doctorat inédite). École Normale Supérieure de Lyon.

Walther, J. B., Slovacek, C. L., & Tidwell, L. C. (2001). Is a picture worth a thousand words? Photographic images in long-term and short-term computer-mediated communication. *Communication Research*, *28*, 105-134.

Wang, Y. (2004). Supporting synchronous distance language learning with desktop videoconferencing. *Language Learning & Technology*, *8*, 90–121.

Wang, Y. (2006). Negotiation of Meaning in Desktop Videoconferencing-Supported Distance Language Learning. *ReCALL: The Journal of EUROCALL*, *18*, 122-146.

Waring, R. (2002). Scales of vocabulary knowledge in second language vocabulary assessment. *Kiyo (Occasional papers of Notre Dame Seishin University, Okayama)*, *26*, 40-54.

Warschauer, M. (1996). Comparing face-to-face and electronic communication in the second language classroom. *CALICO Journal*, *13*, 7-26.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of Reading and Writing on Word Knowledge. *Studies in Second Language Acquisition*, *27*, 33-52.

Webb, S. (2009). The Effects of Receptive and Productive Learning of Word Pairs on Vocabulary Knowledge. *RELC Journal*, *40*, 360-376.

Wesche, M., & Paribakht, T. (1996). Assessing vocabulary knowledge: depth vs. breadth. *Canadian Modern Language Review*, *53*, 13-40.

Wigham, C. R. (2017). A multimodal analysis of lexical explanation sequences in webconferencing-supported language teaching. *Language Learning in Higher Education*, *7*(1), 81-108.

Williams, J. (1999). Learner-Generated Attention to Form. *Language Learning*, *49*, 583-625.

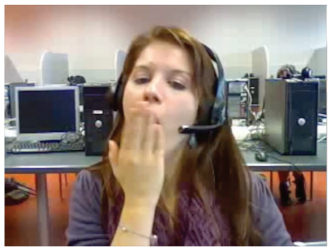
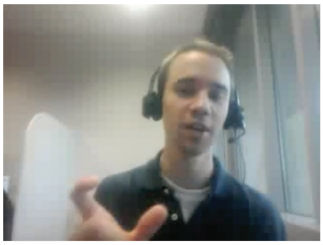
Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). Elan: a professional framework for multimodality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)* (p. 1556-1559).

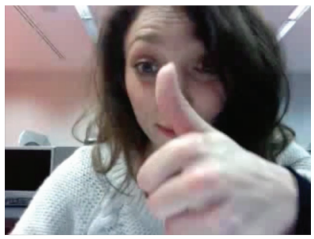
Yanguas, Í. (2010). Oral computer-mediated interaction between L2 learners: it's about time! *Language Learning & Technology*, *14*, 72-93.

Annexes

Annexe 1 : Dimensions des gestes

Ces dimensions de gestes (Ekman & Friesen, 1969 ; McNeill, 1992) ont été utilisés pour classer les gestes dans le cadre de notre mémoire de Master 2 (Holt, 2014).

Geste	Définition	Exemple
Iconique	Le geste iconique illustre des propriétés physiques du référent : taille, forme, action.	 « c'est quand tu fais **(1,1)** »
Métaphorique	Le geste métaphorique illustre un concept abstrait. Les mains servent en général de contenant, tenant le concept clé de la phrase.	 « et donc la différere- **la différence finalement c'est que:** »
Déictique	Le geste déictique est un geste de pointage vers une entité concrète (présente et visible) ou abstraite (absente et invisible).	 « **micro** »

Battement	Le battement est un geste sans contenu sémantique qui rythme la parole. Il peut aussi se superposer à un autre geste, par exemple sur un geste métaphorique pour marquer son importance.	 « on dit qu'il est **paresseux** »
Emblème	Un emblème est un geste qui a une traduction orale exacte (Ekman & Friesen, 1969). Les emblèmes appartiennent à des communautés socio-culturelles et ne sont pas forcément compris d'une communauté à une autre.	 « **MH** »

Annexe 2 : Synthèse non exhaustive d'études réalisées dans le cadre du F1L et de celui de projet ISMAEL

Étude	Participants/ situation	Questions/ objectifs de recherche	Méthodologie	Résultats	Conception de la multimodalité par les auteurs
Develotte, Guichon & Kern (2008) (article de revue)	16 apprentis-enseignants en formation à Lyon enseignent le français à 16 étudiants à l'Université de Californie à Berkeley	Étudier : (1) les effets psycho-affectifs de la plateforme de visioconférence, (2) l'utilisation de la multimodalité, et (3) les compétences développées lors de la formation	Entretiens semi-directifs auprès de l'ensemble des 16 apprentis-enseignants et de 14 sur 16 étudiants américains	Apaisement de stresse et d'anxiété par tous les participants au fil du semestre. Les apprenants deviennent de moins en moins préoccupés par la précision linguistique et favorisent la spontanéité. L'image de l'autre apporte beaucoup à l'échange	La juxtaposition de plusieurs canaux (voix, écriture, image) apporte des avantages didactiques, linguistiques, communicationnels et psycho-affectifs
Guichon & Drissi (2008) (acte de colloque)		Étudier les régulations pédagogiques (consignes, explications, rétroactions et alertes) lors de trois types tâches : échange d'informations, résolution de problèmes et échange d'opinions	Analyse du contenu des tours de parole des tuteurs	Une grille d'évaluation a été élaborée pour mesurer l'efficacité des rétroactions	Seuls le canal oral et le clavardage sont analysés
Guichon (2009) (article de revue)		A la suite des interactions, comment les apprentis-enseignants s'auto-critiquent-ils ?	Analyse du discours des apprentis-enseignants lorsqu'ils s'auto-critiquent lors des débriefings	Les critiques sont catégorisés comme étant descriptifs, l'expression de difficultés, ou des réflexions critiques	Les apprentis-enseignants devraient porter un regard critique sur l'utilisation des outils et des modes, mais ce n'est pas souvent le cas
Dejean-Thircuir, Guichon & Nicolaev (2010) (article de revue)		Quel est le rôle de l'apprenti-enseignant ?	Pour quatre séances, Ils mesurent le taux de participation, la longueur des tours de parole, les silences et chevauchements. Entretiens semi-directifs	Capacités limitées des apprenants. Une variété de pratiques chez les enseignants qui favorisent ou pas l'interaction. L'interaction est facilitée quand il n'y a qu'un apprenant face à un apprenti-enseignant	« quelques indices non verbaux » (382) sont mentionnés. Ils permettent aux apprentis-enseignants de voir la réaction des apprenants et d'ajuster leur comportement
Develotte, Guichon & Vincent (2010) (article de revue)		Comment les apprentis-enseignants utilisent-ils la webcam ?	Étude empirique de l'utilisation des canaux de la plateforme	Trois types de comportement favorisant l'audio, la webcam, et l'alternance entre les deux. Différents degrés d'utilisation de la webcam.	L'interaction entre les modes et la proportion de l'utilisation de chaque mode a des effets sur l'interaction
Guichon, Bétrancourt & Prié (2012) (article de revue)		Quelles sont les stratégies mises en place pour donner du feedback ?	Comptabiliser le nombre de marqueurs posés dans Visu et le nombre d'instances de feedback	Grande variété interindividuelle de l'émission du feedback et de l'utilisation des modes. La maîtrise de ces stratégies est	5 modalités différentes pour donner du feedback : un mode visuelle, trois modes orales, et un mode écrit

				difficile.	
Codreanu & Combe Celik (2012) (article de revue)		Comment les apprentis-enseignants régulent-ils le discours de deux apprenants ?	Analyse interactionnelle et discursive du discours, attention à l'utilisation de la caméra, entretiens	Plusieurs usages et degrés de conscientisation des outils de Visu	La multimodalité peut enrichir les côtés linguistique et socioaffective de l'échange
Codreanu & Combe Celik (2013) (article de revue)		Quels sont les effets de la multimodalité et de la polyfocalisation de l'attention sur l'apprentissage ?	Analyse de l'encadrement de la webcam, utilisation de l'espace, gestes, et politesse des apprentis-enseignants	Différents niveaux d'encadrement de la webcam, différences selon le niveau d'expérience et production de gestes	La disponibilité de plusieurs outils de communication facilite la communication
Guichon (2013) (article de revue)		Quelles sont les méthodes d'analyse ?	Synthèse méthodologique et épistémologique des pratiques réflexives mises en place lors de la formation des enseignants	L'analyse multimodale des interactions en combinaison avec la rétrospection permet aux apprentis-enseignants de s'auto-évaluer et de développer « une conscience sémio-pédagogique critique » (111)	La multimodalité de la visioconférence enrichit l'interaction par rapport aux échanges audiosynchrones
Rivière & Guichon (2014) (article de revue)	18 apprenants dublinois interagissent par visioconférence avec 12 apprentis-enseignants lyonnais	Comment les bilans sont-ils élaborés ?	Analyse discursive des apprentis-enseignants lorsqu'ils construisent le bilan	La création de bilan demande plusieurs compétences diverses	Les canaux audio et visuel permettent de créer des rétroactions enrichies
Holt, Tellier & Guichon (2015) (acte de colloque)		Les apprentis-enseignants gestualisent-ils plus lors des séquences d'incompréhension ?	Comptage de stratégies orales, textuelles et gestuelles lors des séquences normales et séquences d'incompréhension, calcul du taux gestuel	Les apprentis-enseignants ne gestualisent pas plus lors des séquences d'incompréhension	Les apprentis-enseignants disposent d'une palette de modes et doivent sélectionner le mode le mieux approprié à la situation
Guichon & Wigham (2016) (article de revue)		Comment les apprentis-enseignants se positionnent-ils devant la webcam ? Quelles sont les fonctions de gestes ? Quel est le niveau de conscientisation des apprentis-enseignants quant à leur utilisation de la webcam ?	Enregistrement des interactions via des captures d'écran et via une caméra extérieure. Interviews des apprentis-enseignants	Absence d'évolution de l'encadrement de la webcam au fil du semestre. La plupart des gestes sont produits hors champ et sont donc invisibles à l'interlocuteur	
Wigham (2017) (article de revue)		Comment les apprentis-enseignants enseignent-ils le vocabulaire ?	Analyse d'enregistrements vidéo de trois apprentis-enseignants lors de deux séances d'interaction entre Lyon et Dublin	Différentes stratégies multimodales sont employées lors des ouvertures, noyaux et clôtures des explications	Les modes visuels comme les gestes doivent être pris en compte pour analyser l'acte d'enseignement
Thèse de doctorat de Julie Vidal (prévue 2018)		Comment les apprentis enseignants en ligne utilisent-ils les ressources multimodales à leur disposition (et notamment la temporalité) pour fournir une rétroaction	Analyse mixte inspirée de l'ethnométhodologie basée sur l'analyse multimodale des interactions, des bilans asynchrones réalisés par les enseignants et des discours des participants (sous	Différentes stratégies pour ouvrir- fournir et clore une rétroaction corrective et leurs effets sur la co-construction de l'interaction. Mise en valeur de quelques conditions	Les enseignants peuvent tirer profit de la multimodalité pour rendre plus ou moins saillante une rétroaction corrective. Les différentes ressources multimodales utilisées pourraient

		corrective à leurs apprenants ?	forme de bilans réflexifs ou entretiens)	qui rendent une rétroaction corrective mémorable.	avoir un effet sur la mémorabilité de la rétroaction corrective.
Thèse de doctorat d'Aparajita Dey-Plissonneau (prévue 2019-2020)		Quelles sont les affordances programmées et émergeantes dans les interactions multimodales asymétrique pédagogique pour l'apprentissage de L2 à travers la visioconférence ?	Analyse des activités, actions et opérations technologique, linguistique et pédagogique au niveau macro, méso et micro des systèmes d'activité d'apprentissage	Une taxonomie des affordances programmées et émergeantes éducationnelles, technologiques, socio-culturelles et linguistiques. Principes pour le design des outils technologiques et pédagogiques	La combinaison des modes linguistique, paralinguistique et gestuel forme des espaces d'affordances pour réaliser les interactions micro des systèmes d'activité

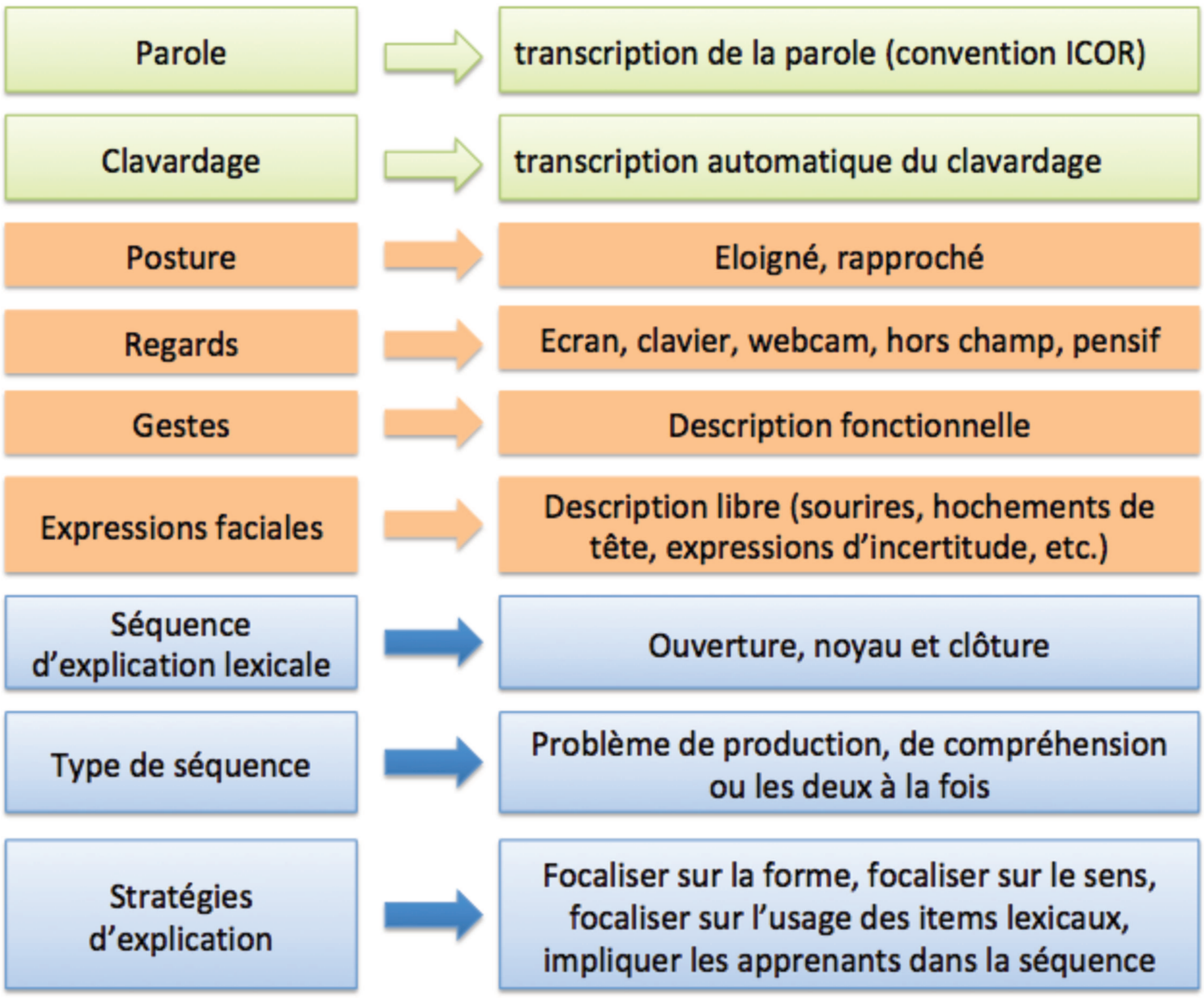
Annexe 3 : Grille d'analyse proposée par Nation & Webb (2011 : 7)

Critère	Note (non=0 ; oui =1)
Motivation	
Y a-t-il un objectif clair lié à l'apprentissage du vocabulaire ?	0 1
Est-ce que l'activité motive l'apprentissage ?	0 1
Les apprenants sélectionnent-ils les mots ?	0 1
Repérage (noticing)	
Est-ce que l'activité dirige l'attention vers les mots-clés /mots cibles ?	0 1
Est-ce que l'activité sensibilise les apprenants à l'apprentissage du vocabulaire ?	0 1
Est-ce que l'activité inclut la négociation ?	0 1
Récupération (retrieval)	
Est-ce que l'activité inclut le la récupération du mot ?	0 1
La récupération est-elle productif ?	0 1
S'agit-il d'un rappel ?	0 1
Y a-t-il des rappels multiples ?	0 1
Y a-t-il du temps entre les rappels ?	0 1
Génération	
Est-ce que l'activité implique usage génératif ?	0 1
Est-ce productif ?	0 1
Y a-t-il un changement marqué qui implique l'usage d'autres mots ?	0 1
Rétention	
Est-ce que l'activité assure que la forme et le sens soient reliés ?	0 1
Est-ce que l'activité implique instanciation ?	0 1
Est-ce que l'activité implique l'imagerie ?	0 1
Est-ce que l'activité évite l'interférence ?	0 1
Score maximal	18

Annexe 4 : Liste des transcripteurs par séance

Enseignant (apprenants)	Semaine n°	Nb. d'interactants	Durée	Transcripteur
Emilie (Fiona, Aiden)	1	3	22m50s	Shuo
Samia (Sean, Angela)	1	3	30m20s	Jean-Guy
Victor (Liam)	1	2	33m25s	Ben
Mélissa (Alejandra, Ana)	1	3	42m11s	Ben
Adèle (Alannah, Catriona)	1	3	43m48s	Ben
Victor (Conor)	2	2	27m46s	Aline
Emilie (Fiona, Aiden)	2	3	33m00s	Lina
Samia (Angela)	2	2	34m34s	Lycia
Mélissa (Alejandra, Ana)	2	3	39m24s	Ben
Adèle (Alannah, Catriona)	2	3	40m55s	Ben
Victor (Liam)	3	2	29m18s	Aline
Samia (Sean, Angela)	3	3	31m50s	Amina
Adèle (Alannah, Catriona)	3	3	33m03s	Julie
Mélissa (Alejandra, Ana)	3	3	37m03s	Thomas
Etienne (Conor, Sophie)	4	3	20m57s	Angèle
Emilie (Fiona, Aiden)	4	3	25m37s	Laurie
Mélissa (Alejandra, Ana)	4	3	30m13s	Angèle
Victor (Liam)	4	2	34m00s	Aline
Adèle (Alannah, Catriona)	4	3	34m27s	Ben
Séverine (Naomi)	4	2	34m47s	Laurie
Samia (Sean, Angela)	4	3	35m20s	Thomas
Victor (Liam)	5	2	29m01s	Aline
Mélissa (Alejandra, Ana)	5	3	32m28s	Thomas
Samia (Sean, Angela)	5	3	39m46s	Lycia
Adèle (Alannah, Catriona)	5	3	41m04s	Thomas
Emilie (Fiona, Aiden)	6	3	24m12s	Jean-Guy
Mélissa (Alejandra, Ana)	6	3	35m53s	Ben
Adèle (Alannah, Catriona)	6	3	35m55s	Ben

Annexe 5 : Schéma d'annotation



Annexe 6 : Tableaux quantitatifs des phénomènes annotés

Annexe 6.1 Adèle

Semaine n°	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Compréhension	7	5	1	1	11	4	**29**
Production	11	5	5	3	9	9	**42**
Les deux			1			2	**3**
Total	18	10	7	4	20	15	**74**

Tableau 1a : Vue globale d'Adèle : les types de pannes lexicales

Semaine n°	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Déclenchée par l'enseignant	8	8	1		9	5	**31**
Déclenchée par l'apprenant	8	1	4	3	4	6	**26**
Déclenchée par les deux	2	1	2	1	7	4	**17**
Total	18	10	7	4	20	15	**74**

Tableau 1b : Vue globale d'Adèle : les initiateurs

Semaine n°	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Forme uniquement	11	7	5		7	7	**37**
Forme et sens	4	2	2	2	8	3	**21**
Forme et usage	2				3	1	**6**
Forme, sens et usage	1	1		2	2	4	**10**
Total	18	10	7	4	20	15	**74**

Tableau 1c : La complexité des explications d'Adèle

Semaine n°	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Microphone	5	2	2	1	3		**13**
Microphone + clavardage	2	5	1	2	6	8	**24**
Microphone + webcam	10	2	2		9	4	**27**
Microphone + clavardage + webcam	1	1	2	1	2	3	**10**
Total	18	10	7	4	20	15	**74**

Tableau 1d : Focalisation sur la forme par Adèle

Semaine n°	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Microphone uniquement	3	2	1		1		**7**
Clavardage uniquement		1					**1**
Webcam uniquement					1	1	**2**
Microphone + webcam	2			4	7	2	**15**
Microphone + Photo					1		**1**
Total	5	3	1	4	10	3	**26**

Tableau 1e : Focalisation sur le sens par Adèle

Semaine n°	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Microphone uniquement	3	1		1	2	3	**10**
Microphone + clavardage				1			**1**
Microphone + webcam					3	1	**4**
Microphone + webcam + clavardage						1	**1**
Total	3	1	0	2	5	5	**16**

Tableau 1f : Focalisation sur l'usage par Adèle

Semaine n°	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Microphone uniquement	1	1			2	2	**6**
Microphone + clavardage						1	**1**
Microphone + webcam	13	6	3	4	10	8	**44**
Microphone + webcam + clavardage	1			1	3	1	**6**
Total	15	7	3	5	15	12	**57**

Tableau 1g : Implication des apprenants par Adèle

Annexe 6.2 Emilie

Semaine n°	**1**	**2**	**4**	**6**	**Total**
Compréhension	7	7	3	3	**20**
Production	4	4	5	5	**18**
Total	11	11	8	8	**38**

Tableau 2a : Vue globale d'Emilie : les types de pannes lexicales

Semaine n°	**1**	**2**	**4**	**6**	**Total**
Déclenchée par l'enseignant	7	7	3		**17**
Déclenchée par l'apprenant	4	1	3	3	**11**
Déclenchée par les deux		3	2	5	**10**
Total	11	11	8	8	**38**

Tableau 2b : Vue globale d'Emilie : les initiateurs

Semaine n°	**1**	**2**	**4**	**6**	**Total**
Forme uniquement	2	5		4	**11**
Forme et sens	5	5	3	3	**16**
Forme et usage	2				**2**
Forme, sens et usage	2	1	2	1	**6**
Aucune focalisation			3		**3**
Total	11	11	8	8	**38**

Tableau 2c : La complexité des explications d'Emilie

Semaine n°	**1**	**2**	**4**	**6**	**Total**
Microphone uniquement	6	11	3	7	**27**
Microphone + clavardage	3		1	1	**5**
Microphone + webcam	1		1		**2**
Microphone + clavardage + webcam	1				**1**
Total	11	11	5	7	**35**

Tableau 2d : Focalisation sur la forme par Emilie

Semaine n°	**1**	**2**	**4**	**6**	**Total**
Microphone uniquement	3	3	2	3	**11**
Clavardage uniquement			1	1	**2**
Microphone + clavardage	2		1		**3**
Microphone + webcam	1	2	1		**4**
Microphone + clavardage + webcam		1			**1**
Microphone + webcam + image	1				**1**
Total	7	6	5	4	**22**

Tableau 2e : Focalisation sur le sens par Emilie

Semaine n°	**1**	**2**	**4**	**6**	**Total**
Microphone uniquement	4	1	2		**7**
Microphone + webcam				1	**1**
Total	4	1	2	1	**8**

Tableau 2f : Focalisation sur l'usage par Emilie

Semaine n°	**1**	**2**	**4**	**6**	**Total**
Microphone uniquement	11	3	3	1	**18**
Microphone + webcam		3	2	1	**6**
Total	11	6	5	2	**24**

Tableau 2g : Implication des apprenants par Emilie

Annexe 6.3 Etienne

Semaine n°	**4**
Compréhension	**3**
Production	**7**
Les deux	**1**
Total	**11**

Tableau 3a : Vue globale d'Etienne : les types de pannes lexicales

Semaine n°	**4**
Déclenchée par l'enseignant	**6**
Déclenchée par l'apprenant	**4**
Déclenchée par les deux	**1**
Total	**11**

Tableau 3b : Vue globale d'Etienne : les initiateurs

Semaine n°	**4**
Forme uniquement	**3**
Forme et sens	**4**
Forme et usage	**1**
Aucune focalisation	**2**
Total	**10**

Tableau 3c : La complexité des explications d'Etienne

Semaine n°	**4**
Microphone uniquement	**5**
Microphone + webcam	**2**
Total	**7**

Tableau 3d : Focalisation sur la forme par Etienne

Semaine n°	**4**
Microphone uniquement	**2**
Microphone + webcam	**2**
Total	**4**

Tableau 3e : Focalisation sur le sens par Etienne

Semaine n°	**4**
Microphone uniquement	**1**
Total	**1**

Tableau 3f : Focalisation sur l'usage par Etienne

Semaine n°	**4**
Microphone uniquement	**5**
Micro + webcam	**3**
Total	**8**

Tableau 3g : Implication des apprenants par Etienne

Annexe 6.4 Mélissa

Semaine n°	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Compréhension	3	1	3	3	8	2	**20**
Production	9	11	12	4	2	8	**46**
Les deux	1						**1**
Total	13	12	15	7	10	10	**67**

Tableau 4a : Vue globale de Mélissa : les types de pannes lexicales

Semaine n°	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Déclenchée par l'enseignant	2	6	5	2	7	3	**25**
Déclenchée par l'apprenant	7	4	9	3	2	5	**30**
Déclenchée par les deux	4	2	1	2	1	2	**12**
Total	13	12	15	7	10	10	**67**

Tableau 4b : Vue globale de Mélissa : les initiateurs

Semaine n°	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Forme uniquement	9	6	8	2	2	6	**33**
Forme et sens	3	4	5	4	6	2	**24**
Forme et usage		1	1			2	**4**
Forme, sens et usage	1		1	1	2		**5**
Aucune focalisation		1					**1**
Total	**13**	**12**	**15**	**7**	**10**	**10**	**67**

Tableau 4c : La complexité des explications de Mélissa

Semaine n°	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Microphone uniquement	5	5	10	2	3	5	**30**
Microphone + clavardage	5	2	3	2	3	4	**19**
Microphone + webcam				2	3		**5**
Microphone + clavardage +	1			1	1		**3**
Total	11	7	13	7	10	9	**57**

Tableau 4d : Focalisation sur la forme par Mélissa

Semaine n°	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Microphone uniquement	3	2		4	3		**12**
Webcam uniquement	1		2				**3**
Microphone + clavardage	1	1					**2**
Microphone + webcam	1	1	2	1	3	1	**9**
Microphone + webcam + image					2		**2**
Webcam + Photo			1				**1**
Total	6	4	5	5	8	1	**29**

Tableau 4e : Focalisation sur le sens par Mélissa

Semaine n°	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Microphone	1	1	1	1		1	**5**
Microphone + webcam			1		2		**3**
Total	1	1	2	1	2	1	**8**

Tableau 4f : Focalisation sur l'usage par Mélissa

Semaine n°	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Microphone uniquement		1	1		1	1	**4**
Webcam uniquement	1				1		**2**
Microphone + clavardage						2	**2**
Microphone + webcam	6	5		3	2	1	**17**
Microphone + webcam +					1	1	**2**
Total	7	6	1	3	5	5	27

Tableau 4g : Implication des apprenants par Mélissa

Annexe 6.5 Samia

Semaine n°	**1**	**2**	**3**	**4**	**5**	**Total**
Compréhension	5	3	1	5	7	**21**
Production	5	5		1	5	**16**
Les deux	1					**1**
Total	11	8	1	6	12	**38**

Tableau 5a : Vue globale de Samia : les types de pannes lexicales

Semaine n°	**1**	**2**	**3**	**4**	**5**	**Total**
Déclenchée par l'enseignant	5	3		4	5	**17**
Déclenchée par l'apprenant	2	4	1	2	4	**13**
Déclenchée par les deux	4	1			3	**8**
Total	11	8	1	6	12	**38**

Tableau 5b : Vue globale de Samia : les initiateurs

Semaine n°	**1**	**2**	**3**	**4**	**5**	**Total**
Forme	3	2			5	**10**
Forme et sens	2	3	1	6	5	**17**
Forme et usage	4	2			1	**7**
Forme, sens et usage	1				1	**2**
Total	10	7	1	6	12	**36**

Tableau 5c : La complexité des explications de Samia

Semaine n°	**1**	**2**	**3**	**4**	**5**	**Total**
Microphone uniquement	6	5	1	3	5	**20**
Microphone + clavardage	5	2		3	7	**17**
Total	11	7	1	6	12	**37**

Tableau 5d : Focalisation sur la forme par Samia

Semaine n°	**1**	**2**	**3**	**4**	**5**	**Total**
Microphone uniquement	2	3	1	4	4	**14**
Webcam uniquement				1		**1**
Microphone + clavardage	1					**1**
Microphone + webcam	1			1		**2**
Microphone + Photo					2	**2**
Total	4	3	1	6	6	**20**

Tableau 5e : Focalisation sur le sens par Samia

Semaine n°	**1**	**2**	**4**	**5**	**Total**
Microphone uniquement	4	1			**5**
Microphone + clavardage		1			**1**
Microphone + webcam				1	**1**
Total	4	2	0	1	**7**

Tableau 5f : Focalisation sur l'usage par Samia

Semaine n°	**1**	**2**	**3**	**4**	**5**	**Total**
Microphone uniquement	5	2	1	4	3	**16**
Clavardage uniquement					1	**1**
Microphone + clavardage	5			1	1	**6**
Microphone + webcam				1		**1**
Microphone + webcam + clavardage					2	**2**
Total	10	2	1	6	7	**26**

Tableau 5g : Implication des apprenants par Samia

Annexe 6.6 Séverine

Semaine n°	**4**
Compréhension	**12**
Production	**7**
Les deux	**2**
Total	**21**

Tableau 6a : Vue globale de Séverine : les types de pannes lexicales

Semaine n°	**4**
Déclenchée par l'enseignant	**13**
Déclenchée par l'apprenant	**5**
Déclenchée par les deux	**3**
Total	**21**

Tableau 6b : Vue globale de Séverine : les initiateurs

Semaine n°	**4**
Forme uniquement	**5**
Forme et sens	**9**
Forme et usage	**2**
Forme, sens et usage	**5**
Total	**21**

Tableau 6c : La complexité des explications de Séverine

Semaine n°	**4**
Microphone uniquement	**12**
Microphone + clavardage	**7**
Microphone + webcam	**1**
Microphone + clavardage +	**1**
Total	**21**

Tableau 6d : Focalisation sur la forme par Séverine

Semaine n°	**4**
Microphone uniquement	**12**
Microphone + clavardage	**2**
Total	**14**

Tableau 6e : Focalisation sur le sens par Séverine

Semaine n°	**4**
Microphone uniquement	**3**
Microphone + clavardage	**1**
Total	**4**

Tableau 6f : Focalisation sur l'usage par Séverine

Semaine n°	**4**
Microphone uniquement	**18**
Microphone + webcam	**1**
Total	**19**

Tableau 6g : Implication des apprenants par Séverine

Annexe 6.7 Victor

Semaine n°	**1**	**2**	**3**	**4**	**5**	**Total**
Compréhension	9	3	5	4	6	27
Production	3	2	4	8	2	19
Total	12	5	9	12	8	46

Tableau 7a : Vue globale de Victor : les types de pannes lexicales

Semaine n°	**1**	**2**	**3**	**4**	**5**	**Total**
Déclenchée par l'enseignant	9	2	5	5	5	**26**
Déclenchée par l'apprenant	2	3	4	1	2	**12**
Déclenchée par les deux	1			6	1	**8**
Total	12	5	9	12	8	**46**

Tableau 7b : Vue globale de Victor : les initiateurs

Semaine n°	**1**	**2**	**3**	**4**	**5**	**Total**
Forme uniquement	3	2	4	7	3	**19**
Forme et sens	7	1	4	2	3	**17**
Forme et usage				2		**2**
Forme, sens et usage	2	2	1	1	2	**8**
Total	12	5	9	12	8	**46**

Tableau 7c : La complexité des explications de Victor

Semaine n°	**1**	**2**	**3**	**4**	**5**	**Total**
Microphone uniquement	11	2	5	6	1	**25**
Microphone + clavardage		1	4	5	5	**15**
Microphone + webcam	1	2		1		**4**
Total	12	5	9	12	6	**44**

Tableau 7d : Focalisation sur la forme par Victor

Semaine n°	**1**	**2**	**3**	**4**	**5**	**Total**
Microphone uniquement	6	2	1	1	3	**13**
Microphone + webcam	3	1	4	1	1	**10**
Microphone + webcam + image			1			**1**
Microphone + image					1	**1**
Total	9	3	6	2	5	**25**

Tableau 7e : Focalisation sur le sens par Victor

Semaine n°	**1**	**2**	**3**	**4**	**5**	**Total**
Microphone uniquement	2	1	1	2	1	**7**
Microphone + clavardage				1	1	**2**
Microphone + webcam		1				**1**
Total	2	2	1	3	2	**10**

Tableau 7f : Focalisation sur l'usage par Victor

Semaine n°	**1**	**2**	**3**	**4**	**5**	**Total**
Microphone uniquement	6		3	1	3	**13**
Microphone + clavardage		1		1		**2**
Microphone + webcam	4	2	1	3		**10**
Microphone + webcam + clavardage				2		**2**
Webcam + image					1	**1**
Total	10	3	4	7	4	**28**

Tableau 7g : Implication des apprenants par Victor

Glossaire

Bricolage lexical : terme employé par Lüdi (1994) pour designer les moyens par lesquels un locuteur non-natif, en situation de détresse verbale, peut « exploiter au maximum la composante lexicale de l'interlangue, voire […] l'étendre au delà de ses frontières actuelles, c'est-à-dire […] construire les mots dont [il] a besoin, avec ou sans l'aide de l'interlocuteur natif » (Lüdi, 1994 : 2)[1]. Le non-natif fait « fait en d'autres termes appel à des techniques de *formulation approximative* ».

Canal : la multicanalité permet la juxtaposition de ressources visuelles (images et texte) et sonores (voix) (Lancien, 1998). Voir *substrat physique*. Les canaux de la plateforme de visioconférence sont la webcam, le microphone, le clavardage, et l'envoi de documents. Les *modes* passent par ceux-ci.

Champ, contre-champ et hors-champ : le *champ* inclut tout ce qui est filmé par la webcam du participant A pour s'afficher sur l'écran du participant B (voir figure ci-dessous). Le champ est modifiable par les participants car ceux-ci « sont libres de manipuler la webcam » (Guichon, 2013 : 103) et de choisir leur façon de se positionner devant celle-ci. Le *contre-champ* permet aux participants de voir leur propre image comme dans un miroir. Cela donne « la possibilité à l'enseignant en ligne d'auto-réguler sa propre image en vérifiant ce qu'il donne à voir à ses apprenants (qualité du cadrage, caractère explicite d'une expression faciale, théâtralisation d'un geste) » (Guichon, 2013 : 104). Enfin, le *hors-champ* inclut « tout ce qui dépasse le cadre de l'écran mais participe à l'interaction pédagogique » : les éléments physiques, comme les meubles et les personnes, ainsi que les éléments symboliques, comme la langue et la culture de l'autre (Guichon, 2013 : 104).

[1] Cette pagination correspond au manuscrit auteur.

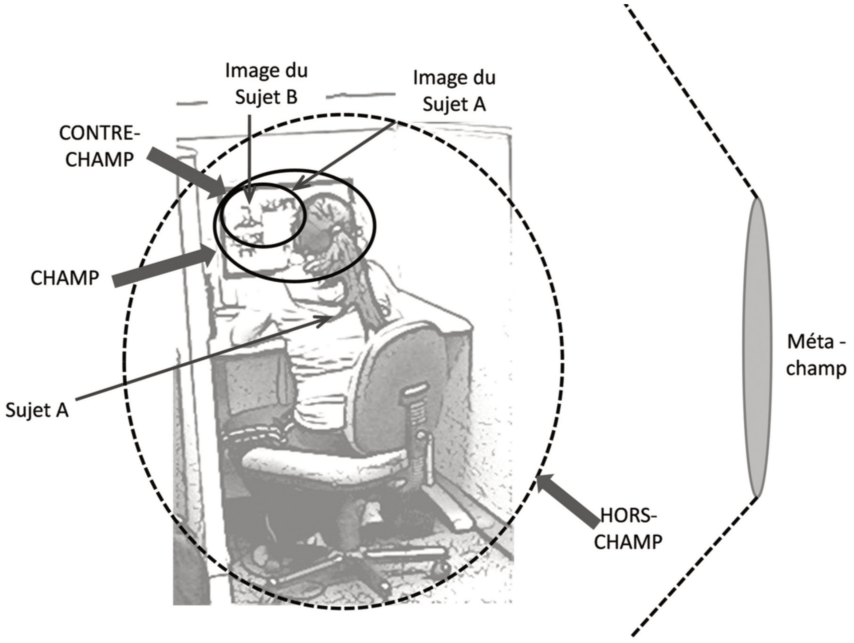


Figure 1 : Les différents points de vue (Guichon, 2013 : 103)

Clôture : voir *Phases d'une séquence d'explication lexicale*.

Densité modale : notion élaborée par Norris (2004, 2014) pour parler de l'hiérarchisation des modes lors de l'interaction. Un mode devient dense lorsqu'il prend une place primaire dans l'interaction ou s'il est entrelacé de manière complexe avec les autres modes.

Double encodage : le cerveau dispose d'un processeur verbal qui traite des mots oraux et écrits, et d'un processeur visuel qui traite des images. Pour l'apprentissage d'un item lexical, on peut utiliser ces deux processeurs à la fois en présentant à l'apprenant une définition verbale et une image en même temps. Les images facilitent le rappel grâce à leur nature concrète (Clark & Paivio, 1991 ; Paivio, 2006).

Glose : signifie courte définition. Une glose orale est donnée immédiatement après l'énonciation d'un mot. Une glose écrite est mise soit entre parenthèses, soit en marge, soit en note de bas de page.

Input : autrement appelé « exposition » en français. L'input est « l'environnement langagier de l'apprenant », « Ce à quoi l'apprenant est exposé » (Cuq, 2003 : 99). L'input ne garantit pas la saisie, ou en anglais l'*uptake* (voir *uptake*). Cuq (2003 : 99) explique que

> « Ce à quoi l'apprenant est exposé peut être constitué d'interactions en face à face, de discours de tous genres, authentiques ou didactiques, sonores ou écrits, et constitue l'apport de données à partir duquel l'apprenant va saisir celles qui l'intéressent, et qui constituent le matériau que traite l'acquisition. »

Médium : l'ensemble des ressources matérielles qui sont utilisées pour produire un évènement sémiotique (Kress & Van Leeuwen, 2001). Par exemple, la visioconférence est un médium qui est constitué de plusieurs ressources matérielles (*substrats physiques*, *canaux*) : la webcam, le microphone, et le clavier.

Mode : la façon dont une ressource sémiotique est formée culturellement par son usage (Jewitt, 2014a ; Kress, 2014). Les gestes et la parole sont deux exemples de modes.

Mot juste : item lexical est parfait pour la phrase ou le contexte en question (Grossmann, 2011), même si on accepte que c'est une illusion car aucun mot et aucune expression ne fait coïncider parfaitement le langage et la réalité.

Négociation du sens : définie par Pica (1994 : 494) comme

> « la modification et restructuration de l'interaction qui a lieu quand les apprenants et leurs interlocuteurs anticipent, perçoivent ou expérimentent des difficultés au niveau de la compréhensibilité du message. En négociant, ils réalisent un travail linguistique afin d'atteindre la compréhensibilité dont ils ont besoin par répétition exacte, ajustement de la syntaxe, changement des mots, ou modification de la forme et du sens par toute autre nature que ce soit »[2].

Noyau : voir *Phases d'une séquence d'explication lexicale*.

Output : terme anglais qui signifie production langagière.

Ouverture : voir *Phases d'une séquence d'explication lexicale*.

Rappel : il y a *rappel réceptif* quand on part de la forme d'un item lexical pour trouver son sens. A l'inverse, il y a *rappel productif* quand on part du sens d'un item lexical pour trouver sa forme. Le rappel productif est plus couteux cognitivement que le rappel réceptif. Si un choix est présenté, il s'agit de la reconnaissance plutôt que du rappel (Nation & Webb, 2011).

Phases d'une séquence d'explication lexicale : nous adoptons un modèle canonique à trois phases (De Gaulmyn, 1991 ; Fasel Lauzon, 2014 ; Gülich, 1990a ; Lüdi, 1991).

1. **Ouverture** : l'item lexical est identifié et problématisé.
2. **Noyau** : l'item lexical est expliqué.
3. **Clôture** : l'enseignant ratifie et évalue la compréhension de l'apprenant.

Phase multimodale : se caractérise par une cohérence entre les différents modes pour une période de temps (Baldry & Thibault, 2006). Lors de la phase d'ouverture d'une séquence d'explication lexicale, tous les modes sélectionnés fonctionnent ensemble pour que l'acteur puisse identifier et problématiser l'item lexical non compris.

[2] « the modification and restructuring of interaction that occurs when learners and their interlocutors anticipate, perceive, or experience difficulties in message comprehensibility. As they negotiate, they work linguistically to achieve the needed comprehensibility, whether repeating a message verbatim, adjusting its syntax, changing its words, or modifying its form and meaning in a host of other ways » (Pica, 1994 : 494).

Principe d'intégration des ressources : lorsqu'on combine un ensemble de ressources sémiotiques, l'effet produit est autre que celui produit par une seule ressource sémiotique. Ce n'est pas la simple somme de chaque ressource (Sindoni, 2013). Par exemple, la phrase « c'est quand tu fait [geste] » serait dénouée de sens sans le geste, et le geste sans la phrase n'aurait pas l'effet communicatif souhaité (cf. chapitre 7 exemple 25 en page 251).

Principe de modalité : l'apprentissage est facilité quand les images et les mots sont présentés ensemble (Fletcher & Tobias, 2005 ; Low & Sweller, 2005 ; Mayer, 2001). Cependant, il est préférable que les mots qui accompagnent l'image soient présentés oralement afin de ne pas surcharger le processeur visuel du cerveau (voir *théorie cognitive*).

Principe de redondance : selon ce principe (Sweller, 2005b), quand on donne plus d'informations aux apprenants, cela nuit à l'apprentissage au lieu de le faciliter. Il y a deux types de redondances :

1) On présente les mêmes informations avec deux modes différents : textuellement et par images ;

2) On présente des informations supplémentaires avec un même mode.

Dans les deux cas, si les informations supplémentaires sont redondantes, la suppression de celles-ci peut faciliter l'apprentissage.

Ressource sémiotique : notion qui englobe l'ensemble des actions, matériaux et artefacts que l'on utilise pour communiquer (Jewitt, 2014a ; Van Leeuwen, 2004).

Séquence d'explication lexicale : une ou plusieurs facettes de la connaissance d'un item lexical sont expliquées. Au sens plus large, les facettes de la connaissance d'un item lexical sont la forme, le sens, et l'usage (Nation, 2013). On les séquences d'explication lexicale en trois phases : l'ouverture, le noyau, et la clôture (voir *Phases d'une séquence d'explication lexicale*).

Substrat physique : terme employé par Bateman (2011) (*physical substrate*) pour parler de la matérialité d'un mode. Par exemple, une feuille de papier est un substrat physique par lequel peuvent passer les modes *texte* et *images*. Les cordes vocales constituent aussi un substrat physique par lequel passe le mode *parole orale*. En visioconférence, la webcam est un substrat physique par lequel peut passer plusieurs modes sémiotiques : gestes, mimiques faciales, regards, postures.

Tâche : selon le CECR (Conseil de l'Europe : 121) « L'exécution d'une tâche par un individu suppose la mise en œuvre stratégique de compétences données, afin de mener à bien un

ensemble d'actions finalisées dans un certain domaine avec un but défini et un produit particulier ». Une tâche peut être simple ou complexe. Selon le CECR, une tâche peut aussi être une activité. Robert (2008 : 194) explique que la tâche « n'est jamais purement langagière, même si elle demande de recourir au langage oral ou écrit, donc à l'emploi d'actes de parole ». Dans notre corpus, les tâches sont les activités planifiées et mises en œuvre par les apprentis-enseignants, par exemple la planification d'un goûter d'anniversaire pour enfants ou l'élaboration d'un menu pour un camion-restaurant. Nous employons donc les termes « tâche » et « activité » de manière interchangeable.

Théorie cognitive de l'apprentissage multimédia : Selon Mayer (2001), l'apprentissage est plus complexe qu'un simple transfert d'informations comme si le cerveau était un récipient vide à remplir. Cette théorie à trois principes de base : (1) les humains disposent d'un canal audio et d'un canal visuel ; (2) ces canaux ont une capacité limitée ; (3) l'apprentissage consiste à sélectionner, organiser et intégrer de nouvelles informations visuelles et auditives. Selon cette théorie, le cerveau possède un processeur auditif et un processeur visuel. Le processeur auditif traite le langage verbal oral, tandis que le processeur visuel traite les images et le texte écrit. Ces deux canaux sensoriels sont indépendants mais peuvent être complémentaires. Si on présente des images avec du texte à côté, le processeur visuel sera surchargé tandis que le processeur auditif ne sera pas utilisé. On peut étendre la capacité cognitive et la mémoire de travail en utilisant les deux processeurs à la fois. Par exemple, quand les informations sont présentées par images avec une narration orale, l'apprentissage est facilité car les deux processeurs sont utilisés ensemble de manière complémentaire (voir *principe de modalité*).

Uptake : Autrement appelé « saisie ». Ce terme

> « désigne, dans un sens restreint, le traitement initial perceptif qu'effectue l'apprenant sur les données linguistiques auxquelles il est exposé, mais tout ce qui est saisi ne devient pas connaissance. Dans un sens plus large, la saisie comprend les processus d'analyse et d'intégration des données dans le système de connaissances de l'apprenant. Elle est généralement considérée comme la phase initiale du processus d'appropriation d'une langue étrangère » (Cuq, 2003 : 218).

L'uptake est donc l'étape qui procède l'input. L'uptake est ce qui est compris, et éventuellement appris par l'apprenant. Or, l'input ne garantit pas l'uptake (Smith, 2005).

Usage génératif : L'utilisation d'un item lexical dans un contexte qui est autre que le contexte dans lequel il a été appris, ou l'utilisation d'une acception d'un item lexical qui est autre que celle initialement apprise (Joe, 1995, 1998).

Table des matières

Remerciements 3
Sommaire 7
Introduction 9
Ancrage 9
Intérêt personnel pour l'apprentissage du vocabulaire 9
Intérêts didactiques et pratiques pour l'entrée par le lexique 11
Objectifs 13
Méthode 13
Plan 14
Partie I : Cadre théorique 17
Chapitre 1 : Le contexte 19
1.1 Introduction 19
1.2 L'enseignement et l'apprentissage des langues médiés par ordinateur 20
1.2.1 L'évolution des TIC 20
1.2.2 Spécificités de l'enseignement par visioconférence 23
1.2.3 Le Français en Première Ligne 25
1.3 Le projet ISMAEL 27
1.3.1 Le dispositif pédagogique 27
1.3.1.1 Les participants 27
1.3.1.2 Le scénario pédagogique 28
1.3.2 Le dispositif technique 28
1.3.2.1 Le salon synchrone de Visu 29
1.3.2.2 Le salon de rétrospection de Visu 30
1.3.3 Le corpus 32
1.3.3.1 Utilité scientifique du corpus 33
1.4 Conclusion 35
Chapitre 2 : Enseignement et apprentissage du vocabulaire en langue étrangère 37
2.1 Introduction 37
2.2 Pourquoi privilégier l'enseignement explicite du vocabulaire en classe de langue étrangère ? 37

2.3 Qu'est-ce qu'un item lexical ? 39
2.3.1 Connaître un item lexical 42
2.4 Qu'est-ce qui favorise l'apprentissage du vocabulaire ? 50
2.4.1 Focalisation sur la forme 51
2.4.2 Focalisation sur le lien entre forme et sens 52
2.4.2.1 Contextualisation 52
2.4.2.2 Groupement sémantique et thématique 53
2.4.2.3 Définitions simples 54
2.4.2.4 Gloses 55
2.4.2.5 Traductions 56
2.4.2.6 Rappel 58
2.4.3 Traitement profond 59
2.4.3.1 Output provoqué par l'enseignant 60
2.4.3.2 Porter un jugement sur un item lexical 60
2.4.3.3 Usage génératif 61
2.4.3.4 Charge d'implication 61
2.4.3.5 Participation active 63
2.4.4 Négociation du sens 65
2.4.5 La tâche 68
2.4.6 Autonomie 70
2.5 Conclusion 71
Chapitre 3 : Vers une approche multimodale de l'analyse des séquences d'explication lexicale en visioconférence 75
3.1 Introduction 75
3.2 Une conception classique des séquences de négociation du sens 76
3.2.1 Modéliser les séquences de négociation du sens 78
3.2.2 Modéliser les séquences d'explication lexicale 81
3.2.3 Rôles des participants 84
3.3 Vers une perspective multimodale 85
3.3.1 Définir la multimodalité 85
3.3.1.1 Ressources sémiotiques 86
3.3.1.2 Mode 87
3.3.1.3 Mode versus médium 87

3.3.1.4 Matérialité des modes 88
3.3.1.5 Définir les modes pour la recherche 89
3.3.2 Les canaux de Visu 89
3.3.3 Les modes sémiotiques 90
3.3.3.1 Parole orale et parole écrite 90
3.3.3.2 Gestes et expressions faciales 91
3.3.3.2.1 Les gestes manuels 93
3.3.3.2.1.1 Gestes et apprentissage du lexique 94
3.3.3.3 Regards 98
3.3.3.4 Proxémique 98
3.3.3.5 Les images et vidéos 99
3.3.4 Synthèse : les ressources sémiotiques 101
3.3.5 Sélection des modes sémiotiques 102
3.3.6 Interaction entre les modes 103
3.3.6.1 Intégration des modes 103
3.3.6.1.1 Les gestes et la parole 103
3.3.6.2 Hiérarchisation des modes 104
3.3.6.3 Phases multimodales 107
3.4 Positionnement épistémologique 108
3.4.1 Intérêt d'étudier les apprentis-enseignants en formation 109
3.5 Conclusion 110
Partie II : Méthodologie 113
Chapitre 4 : Construction du corpus 115
4.1 Introduction 115
4.2 Design 116
4.3 Recueil des données 117
4.3.1 Plans des séances 117
4.3.2 Captures d'écran 117
4.3.3 Caméras extérieures 117
4.3.4 Bilans 117
4.3.5 Débriefings 118
4.3.6 Interviews après le semestre 118
4.3.7 Choix et anonymisation des participants : enjeux éthiques 118

4.4 Construction des vidéos ... 120
4.4.1 Les traces disponibles depuis Visu ... 120
4.4.2 Les traces disponibles depuis Box ... 121
4.4.3 Choix entre les deux façons de recueillir le corpus ... 122
4.4.4 Synchronisation des Flux sur Final Cut Pro ... 123
4.4.5 Montage final sur QuickTime Pro ... 126
4.4.6 Un point de vue « observateur » ... 128
4.5 Anonymisation des images ... 128
4.6 Transcription des données ... 130
4.6.1 Transcription du corpus ... 130
4.6.2 Sélection des séances ... 134
4.7 Partage du corpus ... 135
4.8 Type de corpus ... 136
4.9 Conclusion ... 139
Chapitre 5 : Annotation des données et repérage des séquences d'explication lexicale ... 141
5.1 Introduction ... 141
5.2 Notre place par rapport à nos données ... 141
5.2.1 Notre place par rapport aux enseignants de langues étrangères ... 142
5.2.2 Notre place par rapport aux sujets de notre étude et à nos données ... 143
5.2.3 Notre rapport à la langue ... 144
5.3 Repérage des séquences d'explication lexicale ... 145
5.4 Annotation des phénomènes multimodaux ... 150
5.4.1 Transcription de la parole ... 151
5.4.2 Annotation des ressources visuelles ... 151
5.4.2.1 Postures ... 152
5.4.2.2 Regards ... 152
5.4.2.2.1 Un vocabulaire contrôlé ... 154
5.4.2.3 Expressions faciales ... 154
5.4.2.4 Gestes ... 155
5.4.2.4.1 Espace gestuel ... 156
5.4.2.4.2 Annotation des mouvements ... 158
5.4.2.4.3 Classification des gestes ... 162

5.4.2.5 Documents envoyés ... 164
5.4.2.6 Synthèse : annotation multimodale ... 165
5.4.3 Comptabiliser les annotations ... 166
5.5 Conclusion ... 169
Partie III : Analyses des données ... 173
Chapitre 6 : La focalisation sur la forme des items lexicaux ... 175
6.1 Introduction ... 175
6.2 Focalisation sur la forme : un rapide état de l'art ... 175
6.2.1 Utilité pédagogique de la focalisation sur la forme ... 175
6.2.2 Décontextualiser un item lexical ... 177
6.2.2.1 Les marques voco-verbales ... 179
6.2.2.1.1 Les logonymes ... 179
6.2.2.1.2 Les dislocations ... 182
6.2.2.1.2.1 La dislocation à gauche ... 182
6.2.2.1.2.2 La dislocation à droite ... 183
6.2.2.1.3 La lemmatisation ... 184
6.2.2.1.4 Les pauses et saillance vocalique ... 185
6.2.2.2 Marques multimodales ... 186
6.2.2.2.1 Renforcement par clavardage ... 186
6.2.2.2.2 Expressions faciales ... 187
6.2.2.2.3 Regards ... 188
6.2.2.2.4 Mouvements de tête ... 188
6.2.2.2.5 Les mouvements de bouche exagérés ... 189
6.2.2.2.6 Gestes ... 189
Adèle fait des guillemets gestuels ... 189
6.3 Fonctions didactiques de la focalisation sur la forme ... 190
6.3.1 Problèmes de compréhension ... 190
6.3.1.1 Vérifier la compréhension : une incompréhension anticipée ... 191
6.3.1.1.1 Vérifier la compréhension par autonymie linguistique ... 191
6.3.1.1.2 Vérifier la compréhension avec l'aide du clavardage ... 192
6.3.1.1.3 Vérifier la compréhension implicitement par clavardage ... 192
6.3.1.1.4 Une vérification de compréhension trimodale ... 193
6.3.1.2 Vérifier la compréhension : une incompréhension manifestée ... 194

6.3.1.2.1 La source d'incompréhension n'est pas identifiée 194
6.3.1.2.2 La source d'incompréhension est identifiée 195
6.3.1.3 Résoudre rapidement l'incompréhension d'un mot transparent..... 196
6.3.1.4 Relier la forme au sens 197
6.3.2 Synthèse : problèmes de compréhension 200
6.3.3 Problèmes de production 201
6.3.3.1 Travailler la prononciation 201
6.3.3.2 Travailler l'orthographe 206
6.3.3.3 Augmenter les compétences communicationnelles des apprenants 207
6.3.3.4 La focalisation sur la forme pour distinguer des homophones 208
6.3.3.5 Distinguer des items lexicaux phonétiquement proches 209
6.3.3.6 Donner un mot juste inconnu 210
6.3.4. Synthèse : problèmes de production 212
6.4 Conclusion 212
Chapitre 7 : Aider l'apprenant à accéder au sens d'un item lexical 215
7.1 Introduction 215
7.2 Expliquer le sens d'un item lexical en langue étrangère 215
7.3 Explications voco-verbales 216
7.3.1 Le recours à la langue maternelle 217
7.3.1.1 Usage de l'anglais par les apprenants 217
7.3.1.2 La traduction comme stratégie d'explication 218
7.3.1.2.1 La traduction pour résoudre rapidement l'incompréhension .. 219
7.3.1.3 Attitudes divergentes des apprentis-enseignants à propos du recours à la langue maternelle des apprenants 221
7.3.2 Explications en français 227
7.3.2.1 Définitions générales 228
7.3.2.1.1 Vérifier la compréhension d'un item lexical par l'apprenant 228
7.3.2.1.2 Reprendre une définition de l'apprenant pour donner un mot juste 229
7.3.2.2 Définitions contextualisées par rapport à la tâche 229
7.3.2.3 Synonymes 231

7.3.2.3.1 Synonymes pour résoudre rapidement l'incompréhension d'un mot-clé ... 231
7.3.2.3.2 Confirmer l'hypothèse de l'apprenant et proposer une formulation plus adaptée ... 235
7.4 Explications ostensives ... 238
7.4.1 Usage d'images ... 240
7.4.1.1 Utilisation d'images pour remplacer une explication verbale ... 240
7.4.1.2 Utilisation d'images en combinaison avec d'autres stratégies explicatives ... 243
7.4.2 Usage de sons ... 245
7.4.2.1 Utiliser les affordances de la visioconférence pour transmettre le sens d'un item lexical ... 245
7.4.3 Usage de gestes ... 246
7.4.3.1 Des gestes pour illustrer un concept concret ... 246
7.4.3.2 Des gestes pour illustrer un concept abstrait ... 246
7.4.3.3 Gestes pour vérifier sa propre compréhension et pour décrire le mot juste recherché ... 251
7.4.3.4 Utilisation de gestes pour relier le sens littéral à un sens figuré ... 253
7.5 Conclusion ... 256
Chapitre 8 : Impliquer les apprenants ... 259
8.1 Rappels ... 260
8.1.1 Effectuer un rappel réceptif ... 260
8.1.2 Effectuer un rappel productif ... 265
8.2 Espaces de travail collaboratif sur le lexique ... 268
8.2.1 Faire remarquer un élément problématique ... 268
8.2.2 Faire expliquer à l'apprenti-enseignant ... 273
8.2.3 Faire évaluer une définition ou une proposition d'un mot juste ... 276
8.2.4 Faire répéter la forme correcte ... 279
8.2.5 Faire expliquer à l'autre apprenant ... 280
8.3 Espaces de réflexion sur la langue et sur la culture ... 286
8.3.1 Faire mener une réflexion interlinguistique ou interculturelle ... 286
8.4 Définitions personnalisées ... 289
8.4.1 Personnaliser l'item lexical ... 289

8.4.2 Personnaliser la définition ... 291
8.5 Conclusion ... 295
Chapitre 9 : Expliquer un même item lexical différemment ... 297
9.1 Introduction ... 297
9.2 Bobo ... 301
9.2.1 Opportunité ... 301
9.2.1.1 Transitions naturelles et lisses ... 301
9.2.1.2 Une introduction orientée ... 304
9.2.1.3 Une transition moins fluide ... 305
9.2.2 Complexité et expressivité ... 306
9.2.2.1 Une transition expressive ... 307
9.2.2.2 Une explication lexicographique peu expressive ... 308
9.2.2.3 Une focalisation principalement sur la forme ... 310
9.2.2.4 Une focalisation multimodale la forme et sur l'usage ... 311
9.2.2.5 Une focalisation multimodale sur le sens ... 315
9.2.3 Efficacité ... 319
9.2.4 Synthèse : bobo ... 321
9.3 Goûter d'anniversaire ... 322
9.3.1 Séverine : une revue du lexique ... 323
9.3.1.1 Un espace ménagé pour un travail lexical ... 323
9.3.1.2 Une complexité minimale ... 324
9.3.1.3 Une faible expressivité ... 324
9.3.1.4 Une efficacité immensurable ... 324
9.3.2 Etienne : une explication encadrée par l'activité ... 325
9.3.2.1 Une explication opportune avec un alignement multimodal ... 326
9.3.2.2 Une explication peu expressive orientée vers l'activité ... 326
9.3.3 Samia : une explication simple ... 326
9.3.4 Adèle : une explication riche et expressive ... 328
9.3.4.1 Une explication opportune et complexe ... 333
9.3.4.2 Une explication expressive ... 334
9.3.5 Victor : une séquence d'explication polygérée ... 336
9.3.5.1 Expressivité moindre, implication forte ... 338
9.3.6 Synthèse : goûter d'anniversaire ... 339

9.4 Conclusion ... 341
Chapitre 10 : Discussion ... 343
Introduction ... 343
10.1 Synthèse des résultats et propositions didactiques ... 343
10.1.1 Focalisation sur la forme ... 343
10.1.2 Focalisation sur le sens ... 345
10.2.3 Impliquer les apprenants ... 348
10.2.4 Comparaisons interindividuelles ... 349
10.2 Limites de l'étude ... 351
10.2.1 Focalisation sur les apprenants ... 351
10.2.2 Typologie des tâches ... 353
10.2.3 Efficacité des séquences d'explication lexicale ... 354
10.2.4 Méthodologie et annotation multimodale ... 354
Conclusion ... 356
Conclusion ... 357
Retour sur la démarche entreprise ... 357
Perspectives ... 360
Bibliographie ... 365
Annexes ... 381
Annexe 1 : Dimensions des gestes ... 381
Annexe 2 : Synthèse non exhaustive d'études réalisées dans le cadre du F1L et de celui de projet ISMAEL ... 383
Annexe 3 : Grille d'analyse proposée par Nation & Webb (2011 : 7) ... 386
Annexe 4 : Liste des transcripteurs par séance ... 387
Annexe 5 : Schéma d'annotation ... 388
Annexe 6 : Tableaux quantitatifs des phénomènes annotés ... 389
Annexe 6.1 Adèle ... 389
Annexe 6.2 Emilie ... 391
Annexe 6.3 Etienne ... 393
Annexe 6.4 Mélissa ... 395
Annexe 6.5 Samia ... 397
Annexe 6.6 Séverine ... 399
Annexe 6.7 Victor ... 401

Glossaire 403
Table des matières 409

Séquences d'explication lexicale dans l'enseignement du français par visioconférence : une approche multimodale

Benjamin Holt
Université Lumière Lyon 2, Laboratoire ICAR, financé par LabEx ASLAN
Sous la codirection de Nicolas Guichon et de Marion Tellier

Résumé : Cette thèse vise à étudier l'interaction pédagogique par visioconférence entre sept futurs professeurs de français langue étrangère en formation professionnelle à l'Université de Lyon et douze apprenants de français langue étrangère inscrits dans une école de commerce à Dublin City University. A partir de données écologiques, nous avons construit un corpus audiovisuel de vingt-huit séances et de quinze heures et demie d'interactions. Par une démarche inductive, en nous appuyant sur des méthodes développées dans le cadre de l'analyse conversationnelle, nous avons identifié 295 séquences d'explication lexicale. Nous avons analysé celles-ci par une approche multimodale qui adopte une vue holistique de la communication et prend en compte l'utilisation de l'ensemble des ressources sémiotiques sans privilégier par défaut le langage oral. Nos analyses multimodales montrent de quelles façons les enseignants utilisent les ressources à leur disposition pour se focaliser sur la forme des items lexicaux, pour se focaliser sur leur sens et pour impliquer les apprenants dans les séquences d'explication. Nos données permettent également de mener des comparaisons interindividuelles grâce au fait que plusieurs enseignants, ayant reçu les mêmes consignes et se trouvant dans des situations similaires, expliquent les mêmes items lexicaux.

Mots-clés : visioconférence, multimodalité, explication lexicale, didactique des langues

Lexical explanation sequences in videoconferenced French foreign language teaching: a multimodal approach

Benjamin Holt
University of Lyon 2, ICAR Laboratory, financed by LabEx ASLAN
Under the codirection of Nicolas Guichon and Marion Tellier

Abstract: This thesis aims to study videoconferenced pedagogical interaction between seven future teachers of French who are undergoing professional development at the University of Lyon and twelve learners of French who are enrolled in a business school at Dublin City University. From naturally occurring interaction, we have constructed an audiovisual corpus comprised of twenty-eight sessions and fifteen and a half hours of interaction. Through an inductive approach, and relying on methods developed in the field of conversation analysis, we have identified 295 lexical explanation sequences. We have analyzed them using a multimodal approach which adopts a holistic view of communication and takes into consideration the entire palette of semiotic resources used without automatically prioritizing spoken language. Our multimodal analyses show the ways in which the teacher trainees use the resources at their disposal to focus on the form and meaning of lexical items as well as how they involve their learners in the explanation sequences. Our corpus also allows us to conduct an interindividual comparison due to the fact that several teacher trainees, having received the same instructions and being put in similar situations, explain the same lexical items.

Key words: videoconferencing, multimodality, lexical explanation, foreign language pedagogy