



HAL
open science

A multimodal analysis of lexical explanation sequences in webconferencing-supported language teaching.

Ciara R. Wigham

► **To cite this version:**

Ciara R. Wigham. A multimodal analysis of lexical explanation sequences in webconferencing-supported language teaching. . *Language Learning in Higher Education*, 2017, Synchronous communication technologies in language and intercultural learning and teaching in higher education, 7 (1), pp.81. hal-01452031

HAL Id: hal-01452031

<https://hal.science/hal-01452031v1>

Submitted on 23 May 2017

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.

Wigham, C.R. (2017). A multimodal analysis of lexical explanation sequences in webconferencing supported language teaching. In O'Rourke, B. & Stickler, U. (Eds.) Special issue of *Language Learning in Higher Education: Synchronous communication technologies in language and intercultural learning and teaching in higher education*, 7(1). pp.81-108.

Abstract

Higher Education institutions are increasingly interested in offering more flexible teaching and learning delivery methods that are often independent of place. Where foreign language learning is concerned, telecollaboration is gaining ground. This paper focuses on synchronous webconferencing-supported teaching and examines how different semiotic resources are used during lexical explanation sequences. The context is a telecollaborative exchange between Business students learning French and trainee teachers on a Master's programme in Teaching French as a Foreign Language. Using multimodal transcriptions of interaction data from two sessions, the sequential analysis provides access to different combinations of semiotic resources. These include using the visual mode to project active listening strategies and the complementary role of the text chat to secure common ground concerning the target item. The analysis sheds light on a 'thinking break' strategy employed by the trainees. Descriptive examples demonstrate how verbal explanations were accompanied, firstly, by deictic and iconic gestures and, secondly, by metaphoric gestures used to help forefront different properties of the target item. Finally, changes in gaze and proximity were observed as playing a role in interaction management and in signalling which verbal modality was fronted. The study illustrates emerging pedagogical and multimodal communication strategies for 'doing vocabulary teaching'.

Introduction

A shift has been seen in the Computer-Assisted Language Learning (CALL) field from synchronous written technologies towards audiographic conferencing and both desktop videoconferencing and webconferencing-supported technologies. Language teachers are frequently turning to these technologies to bring together geographically distant learners and trainee teachers in telecollaborative exchange projects. The synchronous nature of these technologies is seen as adding value to their interactions (Levy & Stockwell, 2006). This article contributes to a growing body of research that analyses synchronous interactions from a multimodal perspective, by exploring how trainee teachers during webconferencing-supported interactions allow for a focus on form whilst ensuring that this linguistic focus is context dependent and embedded within the ongoing interaction necessary for task completion. Focusing on the potential of the webcam and its combination with audio and text chat modalities, the study examines how participants manage different semiotic resources during lexical explanation sequences. The paper describes emerging pedagogical and multimodal communication strategies for 'doing vocabulary teaching.'

CALL and multimodality

Guichon & Cohen (2016) argue that whilst any language learning activity is multimodal in nature, in CALL, technology plays a fundamental role in the meaning-making process through "the kinds of meaning which it facilitates or favours, and through the differential access to the means of production and reception which it provides" (Kress & Van Leeuwen, 1996: 233 cited in Guichon & Cohen, 2016:510).

A social semiotic approach to multimodal communication regards the verbal mode as one amongst a multiplicity of modes that may, at different moments in an interaction, take

superior, subordinate or integral roles. The approach is concerned with the choices made by a sign-maker (Halliday, 1978) regarding the combination and interactivity of different communication modes available to present a piece of information within the contextual constraints and potentials of the environment. It considers the social relations within the communication situation and studies the modes available to present information and how this is achieved through different forms of media (Kress, 1993; 2010).

CALL environments may be synchronous or asynchronous. Although the communication modes (e.g. verbal, visual) may be the same in different environments, the nature of the dissemination technology (the medium) and its temporality (asynchronous / synchronous) and nature (static, dynamic, interactive) impact on the types of access that are offered for the presentation and comprehension of information (see Guichon & Cohen, 2016). Within an environment, one mode may correspond to one modality (a specific form of communication related to a single mode), with its own grammar constraining interactions, or one mode can give rise to several modalities. For example, in the verbal mode, information may be communicated in the text chat modality or the audio modality. Consequently, an interaction may be multimodal because several modes are used and/or several modalities (Chanier & Vetter, 2006).

Concerning multimodal environments, Kress considers it vital to "understand the meaning potentials of the resources as precisely and explicitly as we can" (2003:24). This is true in CALL where research needs to assess the ways in which multimodal environments can support the pedagogical processes accepted as effective for language learning (Lamy, 2012) against the limitations on L2 learners' cognitive resources.

Hauck & Youngs (2008) stress that how teachers and material designers can use multimodal CMC settings to exploit communication opportunities and foster SLA is largely unanswered, potentially because only a small body of research examines the impact of the combined use of modes on interaction in online language classrooms. Hampel (2003) describes, in particular, a need to investigate synchronous environments other than written text environments to examine whether the different communication modes can be used by different learner types to their advantage. Lamy, in advocating for CALL research to focus on multimodality, warns that if multimodal meaning resources in L2 studies of computer-mediated communication are not considered "we are in danger of missing out on explaining the nuances in the learning process" (2012b:121) and risk failing to exploit learning processes to the full.

Multimodality and language learning in videoconferencing environments

A small number of studies have been conducted to investigate the potential of online language learning interactions in videoconferencing environments from a multimodal perspective. Research to date has considered how different modes are used in online environments for specific purposes (Guichon & Drissi, 2008; Hampel & Stickler, 2012) or types of discourse (Wang, 2006; Guichon, Betrancourt & Prié, 2012), the effects of the webcam on multimodal interactive learning (Develotte, Guichon & Vincent, 2010; Codreanu & Combe Celik, 2013; Guichon & Wigham, 2016) and social presence (Satar, 2013; Guichon & Cohen, 2014). Task completion and learner proficiency have also been explored (Wang, 2007).

Yamada and Akahori (2009) conducted an experimental study to examine the impact of the visual mode on self-repair by placing learners in different conditions: videoconferencing with both interlocutors' images; with only the partner's image or the learner's own image; without images. They concluded that communication was facilitated when the interlocutor's image was available. This had an effect on learners' perception of social presence and the consciousness of natural communication which led to more self-corrections.

Guichon, Bétrancourt & Prié (2012) examined trainee teachers' feedback provision from a multimodal perspective. The researchers illustrated several strategies across different modalities often based on the trainee teachers' personal preferences and the desire to provide feedback in the text chat in order not to disrupt the communicative flow. The potential of the visual mode to signal incomprehension or the need for self-correction through gestures or facial expressions was barely exploited. The study highlights the need to train teachers in the skills needed to perform in multimodal environments so they do not become cognitively overloaded and in turn reduce feedback provision.

Guichon & Wigham (2016) examined the meaning-making potential of the webcam from a semiotic perspective by exploring how trainee teachers use its affordances to produce visual cues (communicative and extra-communicative gestures, proxemics) useful for mutual comprehension. The study found, firstly, that whilst a head and shoulders framing shot was favoured by the trainee teachers, there did not appear to be an optimal framing choice for desktop videoconferencing. Secondly, there was a loss between the number of gestures performed and those visible to the students. Thirdly, when audio and kinesic modalities were combined, communicative gestures that were framed, and held long enough to be perceived by the learners, were more likely to be valuable for mutual comprehension.

Although the above studies highlight the potentials of multimodal aspects of videoconferencing for language learning, the use of this synchronous technology does not come without challenges. With reference to online language learning and teaching contexts, Wang (2006) warns that the multimodality can put pressure and strain on users and Guichon *et al.* (2012) underline the difficulties of handling the interaction across several modalities, particularly regarding 'dual tasking' where two tasks in different modalities were put in competition in terms of the teachers' cognitive resources.

Lexical explanations during task-based interactions

One challenging aspect of online pedagogy that teachers face during task-based interactions in synchronous environments is how to extract new linguistic material, including lexical items, to allow for a focus on form whilst ensuring this focus is context-dependent and embedded within the ongoing interaction necessary for task completion. Mortensen (2011) describes how a spontaneous focus on lexical material can be analysed, from a conversation analysis perspective, in terms of repair (Schegloff, Jefferson & Sacks, 1977), whereby the participants deal with a problem as it appears in the interaction. Either the teacher identifies a lexical item in his/her verbal turn as potentially problematic and provides an explanation, sometimes after asking the learners explicitly to state their (in)comprehension, or the lexical item is located and pointed out by the learners as problematic (other-initiated repair). Mortensen terms such sequences as "doing (vocabulary) teaching" (2011:136) to frame the pedagogical aspect as a social practice during which

lexical explanations are constructed in and through the interaction, only diverting from the latter for a short 'side-sequence' (Fasel Lauzon, 2008:83) within the ongoing task accomplishment.

In the first instance, the teacher either tests the learner's lexical knowledge by asking a known-information question¹ to determine whether he is able to display his knowledge of the lexical item, or explicitly asks the learner whether he understands the lexical item. This leads either to a learner explanation to which the teacher may react; a signal of comprehension to which the teacher may follow up on by asking a display question; or a signal of incomprehension that prompts a teacher explanation to which, in turn, the learner reacts.

The prototypical teaching sequence that is associated with other-initiated repair, when a learner does not understand an element in the teacher's turn, is often described by a ternary model (Gulich, 1990) in which a lexical difficulty is manifested either explicitly or implicitly by the learner and managed by the teacher through an explanatory reaction to resolve the problem, which in turn the learner ratifies.

Although these models offer a clear starting point for examining lexical explanations in pedagogical contexts, real-time interaction between learners and teachers may not be as clear as these models indicate because when "discourse is stripped down to its lowest common denominator, that is, that of text [...] the 'reality' of the discourse situation is lost during this process" (Knight, 2011:2) including the multimodal properties of the interaction. A multimodal analysis of lexical explanations may reveal aspects of the learning sequence that would not appear if examining the verbal mode alone and may help to better understand how trainee teachers interact with and use the multimodal constituents of webconferencing-based platforms to identify strategies that could be explicitly taught during teacher training.

Multimodal analyses of lexical explanations

Several studies from the Gesture Studies field conducted in classroom environments and under experimental conditions have included a multimodal perspective while examining foreign language teachers' lexical explanations. The research suggests that learners receive considerable input in nonverbal form that is salient for second language acquisition because it modifies and makes verbal input more comprehensible (Lazaraton, 2004). Faraco & Kida (2008) illustrated how facial signs can project lexical difficulties and make visual the different stages of the cognitive activity of the learner confronting difficulty whilst Tellier & Stam (2010) show that illustrative iconic gestures to clarify and disambiguate meaning are more often used with language-learners than with native speakers.

However, in CALL, few studies have examined spontaneous lexical explanations in webconferencing-based interactions from a multimodal perspective. They have examined lexical explanations within the wider context of instructional regulations that also included instructions and feedback (Guichon & Drissi, 2008) or peer scaffolding (Cappellini, 2013).

¹ Also termed display questions. These are typically questions to which the questioner already knows the answer.

Guichon & Drissi (2008) conducted a content analysis of interactions between trainee teachers of French and American learners of French that took place over an eight-week period via a desktop videoconferencing platform. They identified different pedagogical strategies used by the trainees. Whilst some immediately offered lexical explanations, others used inference strategies that included eliciting examples, peer explanations, or associating the text chat and audio modalities to facilitate understanding.

Cappellini (2013) examined a teletandem exchange between Chinese students of French and Master's students at the Language faculty of a French University. His qualitative study underlined different visual strategies during the manifestation phase, including changes in posture and head movements alongside facial mimics of incomprehension. Also, the use of the text chat modality appeared only partway through the interactions (fourth session) during the core phase of explanations.

In line with these two studies, this paper proposes a "multimodal analysis" of one aspect of synchronous online pedagogy and explores the orchestration of different semiotic resources during both during trainee-initiated and learner-initiated (unplanned) lexical explanations.

Methodology

Pedagogical Context, Participants and Ethics

This study draws on interaction data from a telecollaborative exchange involving 18 Business undergraduates in their third semester of learning French at Dublin City University and 12 trainee teachers on a Master's programme in Teaching French as a Foreign Language at Université Lyon 2. For the Master's students, the exchange formed part of an optional module in online teaching. For the Dublin City University undergraduates, the exchange was one element of a 12-week blended French for Business module with level B1.2 of the Common European Framework of Reference for Languages as its minimum exit level (Council of Europe, 2001). The exchange was aligned with module outcomes that included understanding spoken language on topics related to professions, work placements and job applications and applying for an internship through French. This was pursued by most learners six months after the exchange.

The six-week telecollaborative exchange ran in autumn 2013. Participants met for weekly 40-minute online sessions in the webconferencing platform *Visu* (Bétrancourt, Guichon & Prié, 2011). Following an introductory session, each online session was designed by two trainees and was thematic, focusing on Business French. Topics included preparing for an internship, project management, pitching a project, interviews, labour law. Although this paper focuses on the synchronous sessions, these interactions were part of a larger circular learning design that included debriefing sessions where trainees shared extracts from past interactions for collective exploration, and multimodal reports that offered asynchronous feedback to the learners after each session (detailed in Guichon & Wigham, 2016).

Groups were formed randomly by both course lecturers. Gender, teaching experience, and language level were not taken into account. Because of differently sized groups, most trainees worked with two learners whilst two trainees had one-to-one sessions.

This study focuses on interaction data from the third and fourth synchronous sessions of the exchange for three of the exchange groups. The topic for session three was ‘preparing for an internship in Reims.’ Participants engaged in an interview role play for an internship at a town council related to events marketing. Session four centred upon project management. Learners took on the role of a fast-food chain’s employees who were meeting their manager to suggest new ideas for a children’s birthday party formula.

The *Visu* platform was chosen for the web-conferencing interactions because it was specifically designed for synchronous language teaching-learning. Interactions were influenced by its functionalities (Table 1).

Audio	Full duplex (several speakers can speak simultaneously)
Text chat	No private channels available
Video images (via webcam)	Large pictures of interlocutors and thumbnail images of individual
Markers	Personal notes that can be placed on the session timeline by trainees
Multimedia activity resources	Notes, images, videos, hypermedia links that can be pre-prepared by the trainees before the sessions
Automatic recording	Automatic recording of sessions (allowing trainees later re-viewing for training purposes)

Table 1: Functionalities of *Visu*

Participants connected to *Visu* from university language labs using individual computers and headsets. Learners working in the same group were physically separated in the lab.

Participation in the research study was voluntary: not all participants involved in the pedagogical exchange gave permission. Prior to data collection, approval from Dublin City University’s ethics committee was gained. To preserve anonymity, participants’ names have been changed and, when requested, participants’ images have been converted to line drawings.

Research Design

The ISMAEL corpus (Guichon *et al.* 2014) comprises the interaction data of six sessions for seven trainee-learner groups. This study uses data from sessions three and four for three of the exchange groups (Table 2). The rationale for selecting these sessions was that the trainees had gained some online teaching experience and established social relationships with the learners. At the time, the lengthy process of extracting the webcam videos from the *Visu* database and aligning them in one video file that could be imported into the transcription software ELAN (Sloetjes & Wittenburg 2008) had been completed for these

three groups. The spoken interaction had also been transcribed using the conventions given in Appendix A. Using the timestamps created in *Visu*, the parallel text chat logs were synchronized with these transcriptions.

Trainee		Victor	Samia	Adèle
Learners		Liam	Sean & Angela	Alannah & Catriona
Session length and number of verbal turns	Session 3	29min18s 656	31min55s 177	33min03s 1018
	Session 4	34min00s 958	35min21s 783	34min28 2116

Table 2: Data extent

The first step in the analysis process was to identify lexical explanations in the data. This collection was identified in two ways: lexical explanations that followed display questions (e.g. *can you explain what x means?*) or questions to evaluate comprehension (*do you understand the word x?*) and had therefore been anticipated as potentially problematic by the trainees, and unplanned explanations that followed a learner's affirmation that he had encountered a comprehension problem as shown through prototypic phrases (e.g. *what is x? what does x mean?*). The sequences were isolated from the trainee or learner question until the turn in which a learner either asserted verbally or co-verbally his understanding/reception or in which a transition to a new activity was observed. Across the six sessions studied, there were 14 lexical explanation sequences in total of which only four were initiated by the learners.

Secondly, co-verbal acts used in each lexical explanation sequence were annotated in ELAN. In the data, these included gestures, actions, proxemics, head movements and changes in gaze.

McNeill's (1992) schema that categorises gestures as iconic (representations of an action or object), metaphoric (illustrating an abstract concept), deictic (pointing gestures at concrete or abstract spaces) and beats (movements to accompany the rhythm of the discourse) was used to annotate gestures. Kendon's 1982 category of emblems, referring to culturally specific gestures was also included. Actions that contributed to the communication, such as writing something down or typing, were also annotated, but extra-communicative actions such as a participant pushing his hair behind his ear were not.

To annotate head movements, the distinctions offered by Altorfer *et al.* (2000) between rotational (shaking the head), lateral (tilting the head) and sagittal (nodding) movements

were used. Changes in proxemics, when a participant moved closer to or further away from the webcam were annotated. Gaze was annotated as being directed towards the screen, towards the keyboard, outside of the webcam frame and pensive (outside of the webcam frame with gaze directed upwards). Moments when participants' eyes were closed were also annotated.

Analysis

The following sequential analysis offers descriptive illustrations of different strategies used during the opening, explanation and closing phases of student-and teacher-initiated lexical explanation sequences.

Trainees' strategies to highlight the target lexical item

Trainee-initiated sequences (TIS)	Freq.
vous savez/ tu sais ce que c'est X <i>you know what X is</i>	3
est-ce que vous savez / tu sais ce que c'est X <i>do you know what X is</i>	3
tu connais X <i>do you know X</i>	1
et X c'est quoi alors Student_name/ <i>and X so what is it Student_name/</i>	1
est-ce que tu connais + description/ <i>do you know + description/</i>	1
est-ce que tu connais l'expression X <i>do you know the expression X</i>	1

Table 3: Types of teacher-initiated sequences

In trainee-initiated sequences, trainees' turns are formulated to frame the target lexical item (Table 3). The display questions that open the sequences follow a pause and comprise a change in direction of an ongoing turn constructional unit containing the target item (examples 1-2).

Example 1 Samia (S3, 20:15-20:20)

et Angela/ + sera ton manager\ + vous savez c'que c'est un manager\

and Angela/ + will be your manager\ + you know what a manager is

Example 2 Victor (S4, 04:26-04:34)

on va réfléchir à une nouvelle formule pour les goûters d'anniversaires + est-ce-que tu sais ce que c'est un goûter d'anniversaire/

we are going to think about a new formula for the birthday parties + do you know what a birthday party is/

In eight of the ten turns that comprised the openings of trainee-initiated sequences, the target lexical item occurred in a possible turn constructional unit final position (TCU, Sacks *et al.*, 1974) where turn-transfer from the trainee to the learner(s) is possible (Table 3). The trainees' turns are completed by the target item and followed by a pause that shows the possible completion of the turn constructional unit and offers learner(s) a turn-taking opportunity. Prosody appears to play a role in the highlighting the lexical item in its TCU final position – in all cases, the trainees either stress the target item, elongate the pauses between the individual words in the item or, compared to the preceding turns and even the question form itself, produce the item at a slower speed.

Trainees also combined the audio and written modalities by providing the lexical item in the text chat (5 instances) or drawing attention to its written form in resources the learner had already opened (1 instance). This strategy was accompanied by gaze changes (see Figure 1). During the first TCU when the target item was pronounced, the trainee, Samia, shifts her gaze towards the keyboard, perhaps signalling that the pause before the display question comprised a change in direction of an ongoing turn rather than a transition relevance place² (Figure 1, plate 2). When the text chat message appears, the trainee changes her gaze (Figure 1, plate 3), seemingly towards the message on the screen before completing her audio turn. In the pause following the display questions (Figure 1, plate 4), the trainee smiles, to signal the end of her turn and encourage her learners to take the floor. The trainee's gaze then changes towards the other side of the screen (Figure 1, plate 5) presumably towards the learners' webcam images. This gaze change could be to monitor the visual mode for cues that the learners understand the target item or could help signal her wish for speaker transition.

² Transition relevance places are junctures in interaction where transition to another speaker is a possible next action. The transition may, but need not, occur (see Sacks, Schegloff and Jefferson, 1974).

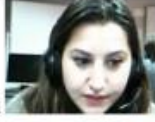

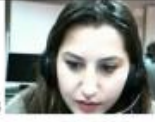
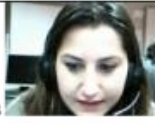
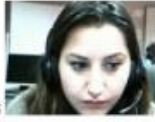
Samia 21:46- 21:51	1 	et vous allez construire ce qu'on	and you are going to make what we
	2 	appelle un pense-bête + vous savez ce que c'est un	call a reminder, +do you know what it is
Text chat (Samia) 21:51		un pense bête	a reminder
Samia 21:51- 21:51	3 	un pense-bête	a reminder
Silence 21:51- 21:53	4 		
	5 		

Figure 1: Text chat modality and gaze changes

In this example, the text chat modality provides a complementary role to the audio modality. Should there be any technical issues, the text chat allows participants to establish common ground concerning the lexical item in question. It also allows the trainee to model the language in writing and therefore provide the spelling of the item, in addition to the target pronunciation. As text chat is a less ephemeral modality (the message remains on the screen), this strategy allows the learners to return to what is written.

Trainee's active listenership strategies in the openings of learner-initiated sequences

In the four learner-initiated sequences, the learners explicitly state their incomprehension. In Figure 2, the learner-initiated sequence follows the trainee visually displaying the instructions and asking whether they understand the activity (*does that make sense to you?*) before telling them to start the interaction (*okay I'm listening to you*). The learner, Alannah, who asks for an explanation has her hands over her earphones. This is a posture she often adopts during the interactions, rather than a co-verbal gesture to suggest that this specific incomprehension of the target item stems from technical problems. Indeed the sequence opening is prompted by a visual display of the target item rather than the item being used in the audio modality as is the case in Figure 3. In this example, the learner, Sean, thins his lips just after the trainee's pronunciation of *le bilan* (the reminder) (Figure 3, plate 1) and changes posture, moving closer to the webcam during his clarification request (Figure 3, plate 2). These signs could be projective of his difficulty. Bavelas and Chovil (1997) describe, for example, how pressing lips together is a back-channel signal to show concern.

What is perceptible in the openings of the learner-initiated sequences is how trainees visually show their engaged response during learner requests. In Figure 2, the trainee moves

closer to the webcam and shifts gaze to the left-hand side of her screen, presumably to her learner's webcam image. In Figure 3, the trainee moves slightly closer to the screen and smiles just after the learner pronounces the problematic lexical item. These active listenership strategies demonstrate trainees' responsive role in the interaction (see Tsuchiya, 2013). They could be considered as proxemic engaged-response tokens: rather than using verbal response tokens (e.g. mmm), the trainees foreground the visual mode to show their availability to deal with learner requests.

Alannah 10:36- 10:39	1 	qu'est-ce que c'est euh les goûter "is it like	<i>what is uh the party + is it like</i>
	2 	the taste"/ ["flav-"/]	<i>the taste/ ["flav-"/]</i>

Figure 2: Changes in screen proximity

Samia 02:05- 02:07	1 	si non tu as vu le bilan ou pas que j'ai préparé pour toi/	<i>otherwise did you see the recap or not that I prepared for you/</i>
Sean 02:10- 02:10	2 	bilan/ c'est quoi /	<i>recap/ what's that/</i>

Figure 3: Active listenership

Setting up the explanations in learner-initiated sequences

Learner-initiated sequences	Freq.
X c'est quoi/ <i>X what is it/</i>	1
qu'est-ce-que c'est X/	2

<i>what is X/</i>	
qu'est-ce que c'est X is it like Xtranslation/	1
<i>what is X « is it like Xtranslation/ »</i>	

Table 4: Types of learner-initiated sequences

Four learner-initiated lexical explanation sequences were identified (Table 4). Before trainees begin an explanation they systematically check the target word in the learner's turn and often give themselves a short thinking break before proceeding.

Foregrounding the text chat was a strategy used to check the target lexical item (Figure 4). The trainee repeats the problematic item (Figure 4, plate 1), and begins to contextualize it before repeating the beginning of the composite noun. The learner, Alannah, expresses her understanding of *goûter* as being the homonym to the infinitive verb form through a direct translation (*taste*) and an acknowledgement that *goûter d'anniversaire* is the item causing difficulty. The trainee, Adèle, then announces a change as to which modality is foregrounded. A gaze and posture change (Figure 4, plate 3) contextualise the deictic (*here*) in the audio. This change prompts Alannah's gaze to move downwards presumably to read the text chat message (plural noun form *goûters*), demonstrating understanding that it is the noun form rather than the verb form that is unclear.



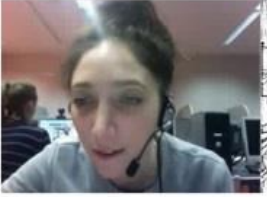



Adèle 10:40- 10:43 1			[euh goûter/] + gouter d'a -[euh goûter/]	[uh party/] + birthday par / [uh party/]
Alannah 10:42- 10:43 2			["taste"/] yeah	["taste"/] yeah
Adèle 10:44- 10:48 3			alors+ euh goûter:/ + hop + attends \ + ce mot là/	so + uh party/ + ah hang on \ + this word here/

Figure 4: Checking target word

Recurrently, once the participants confirmed the target word for which a learner had requested an explanation or the learners had shown their incomprehension of a vocabulary item following a display question, the trainees adopted thinking strategies before proceeding with the explanation. These strategies were communicated visually and verbally.

For example, after a confirmation check (Figure 5, plate 1), the trainee repeats the target item before a shift in gaze off-screen that is preceded by the hesitator *uh*:. Goodwin and Goodwin (1986) describe, in L1 face-to-face interaction, that when a speaker displays hesitancy in a word-search episode, a prototypical action is to accompany a sound stretch with withdrawing gaze from the recipient and a characteristic ‘thinking face’ described by Bavelas et al. (2014) as a break in eye contact and an effortful facial expression. Here the trainee displays hesitancy in his utterance-formulation and difficulty in finding the relevant items to explain the lexical item. This action may aid the trainee to maintain a sense of coherence for himself and help in his thinking process (Codeanu & Combe-Celik, 2013) but will not go unnoticed during the interaction (see Guichon & Wigham, 2016), particularly if the thinking strategy involves changes in gaze and posture accompanied by gestures (Figure 6). In this example, the trainee is present on screen in the visual mode but momentarily absent in the interactional space as she changes her posture and gaze towards somebody who does not appear in the webcam view. During her thinking-out-loud process (Figure 6, plate 1) an off-screen interaction appears to begin in the face-to-face interactional space. The question (*how can I explain birthday party*) appears to be directed not to the trainee herself or her learners but to a peer. The trainee then performs a body-focused gesture (Figure 6, plates 2 & 3) prompting one learner (Alannah) to state her understanding of the base form (*goûter*) thus offering a starting point for the explanation. The trainee gaze shifts and shows visually that she is again present in the online interactional order.




Victor 07:11- 07:11 1			qu'est-ce que c'est quoi !	<i>what's what/</i>
Sean 07:12- 07:13 2			des liens	<i>links</i>
Victor 07:13- 07:14 3			ah des liens/ c'est euh:	<i>ah links/ it's uh:</i>

Figure 5: Thinking face







Adèle 10:50- 10:55	1 		c'est euh: + goûter d'anniversaire \ comment est-ce que je peux	<i>it's uh + a birthday</i> <i>party</i> \ <i>how can I</i>
	2 		expliquer goûter [(rire) un goûter:]	<i>explain party/</i> <i>[(laughs) a party]</i>
Alannah 10:54- 10:55	3 		["cause it's like"] je sais que c'est euh: taste / [ma]is\	<i>["cause it's like"] I</i> <i>know that it's like</i> <i>uhtaste [bu]t</i>

Figure 6: Visually thinking out-loud

Multimodal explanations

In the audio modality, trainees employed different strategies to define the lexical items (Table 5). For the trainee-initiated sequences, the co-elaboration of explanations was privileged. The data included three examples of substitution whereby a lexical item (synonym) or paraphrase was offered.

Learner-initiated sequences (SIS)	Freq.	Trainee-initiated sequences (TIS)	Freq.
substitution	1	co-elaboration (learner suggestion → refinement)	4
translation + contextualisation	1	substitution	3
hypernym + translation	1	formal definition	2
formal definition	1	contextualisation	1

Table 5: Verbal strategies

The verbal strategies used for lexical explanations in the data did not stand alone and were only part of what learners were presented with. Combination with the visual mode allowed trainees to use gestures to add important information – for example, to help contextualise the lexical item by using a situation to illustrate the problematic item, as demonstrated by Victor's explanation of 'retroplanning' (Figure 7).

Firstly, in the verbal mode the trainee suggests that, when you plan an event, you start from one point. He combines the verbal and visual modes to give a suggestion (*today*) and an abstract deictic gesture points towards *today* that the trainee wants to be a point of

common understanding (Figure 4, plate 3). The trainee then demonstrates the pointing gesture moving to the abstract reference to which he points in Figure 4, plate 4 (*the date of the event*). Presuming that this reference is secured, Victor then reuses a deictic towards this reference, accompanied by *the date of the event* and explains through a moving gesture that you plan backwards.













Victor 15:13- 15:29			c'est quand: euh: tu planifies un évènement/ mais euh: en partant/	<i>it's when: uh: you plan an event/ but uh: starting/</i>
1			au lieu de partir par exemple	<i>instead of starting for example</i>
2			d'aujourd'hui ah: jusqu'à la date euh:	<i>from today ah: until the date uh:</i>
3			de l'évènement/ tu fais l'inverse tu pars de:	<i>of the event/ you do the opposite you start from:</i>
4			la date de l'évènement et tu fais ta planification	<i>the date of the event and you do your planning</i>
5			à rebours	<i>backwards</i>
6				

Figure 7: Deictic gestures during a contextualization

Victor proceeds by giving a further contextualised example (Figure 8). In the second contextualisation, the trainee reuses the common referent of an activity that needs to be planned through the use of deictic gestures that contextualise his utterances in the verbal

mode (*this this and this*). However, these gestures are not clearly framed in the webcam image and are barely accessible to his learner (Figure 8, plate 2). The trainee restates that you need to plan backwards and reuses a pointing gesture towards the gestural space for which the reference was previously secured as referring to the date of the final event (Figure 7, plate 5 & Figure 8, plate 4) and a diagonal gesture expressed with *backwards* (Figure 8, plate 5). In this example, the deictic gesture is no longer used with a verbal reference to *the date of the event* as this is understood as having been secured (Figure 8, plate 4). Rather, this gesture gives extra information about where you should start. The trainee does not need to explicitly restate that this is the date of the event as this information is communicated through his abstract deictic gesture. This gesture adds important information to the second contextualisation that is not given in the verbal mode, and cohesion to the explanation by tying together thematically related parts of the verbal discourse that are temporarily separated within the explanation.


Victor 15:31- 15:55		<p>donc par exemple si j'organise quelque quelque chose pour le 25 décembre + pour Noël+ je décide euh je me dis bon ça me prends les deux derniers jours + je dois préparer ça</p>	<p><i>so for example if I organise some something for the 25th of December + for Christmas ° I decide uh I say to myself okay that's going to take the two last days + I need to prepare this</i></p>
1		<p>ça ça + euh : 10 jours avant</p>	<p><i>this this + uh: ten days beforehand</i></p>
2		<p>je dois avoir acheté tous mes cadeaux / etc etc tu fais une planification en</p>	<p><i>I need to have bought all of my presents/ etc etc you're planning</i></p>
3		<p>partant en faisant</p>	<p><i>starting going</i></p>
4		<p>le chemin inverse finalement</p>	<p><i>backwards in fact</i></p>
5			

Figure 8: Gesturing to previously secured references

The trainees also used iconic gestures, to represent an action or an object, in their lexical explanations. For example, in Figure 9, when explaining birthday party, the trainee code switches and tries to use the English (*cake*). This is accompanied by a questioning facial expression (Figure 9, plate 1), showing to her learners that she may not be very sure of her translation. She then tries to explain using a context but, realising this may be culturally specific, adopts a questioning intonation throughout her turn. Adèle proceeds by using a numeric gesture (*four o'clock*) to explain that children eat. She mimes this action with an iconic gesture illustrating the physical movement of eating (Figure 9, plate 3) before using another gesture that accompanies 'something' and illustrates the physical property of the item she is referring to as small. The trainee then checks learners' comprehension of her explanation. This is not done verbally. Instead, the trainee uses the visual mode to seek a response through a questioning gesture (both hands, palm upwards, Figure 9, plate 6) accompanied by a facial expression and directs her gaze towards the webcam images of the learners. These visual cues signal a turn shift and offer learners the possibility to take the floor.

<p>Adèle 11:14- 11:23</p> <p>1</p>		<p>cake vous savez les en- fants/ + peut-être ils font en irlande cake + aussi quand il est euh + euh</p>	<p><i>cake you know? chi- children/ + maybe they do this in Ireland too when it's uh + uh</i></p>
<p>2</p>		<p>quatre heures/ m- milieu de la journée/ + [euh]</p>	<p><i>four o'clock/ m- middle of the day/ + [uh]</i></p>
<p>Alannah 11:23- 11:25</p> <p>3</p>		<p>[et euh je sais euh]</p>	<p><i>[et erm I know uh]</i></p>
<p>Adèle 11:23- 11:25</p> <p>4</p> <p>5</p>		<p>[on mange un: un un]</p> <p>un truc</p>	<p><i>[we eat a a a]</i></p> <p>a snack</p>
<p>Silence 11:25- 11:27</p> <p>6</p>			

Figure 9: Use of iconic gestures

A further multimodal strategy used by the trainees in their explanations was to use the text chat to send a hyperlink to an image. Figure 10 illustrates this strategy. The trainee, Victor, is explaining the lexical item 'mind map' (*carte mentale*). The sequence was trainee-initiated and occurred during the instruction-giving stage of an activity. As discussed previously, the trainee uses both a verbal and visual hesitator (Figure 10, plate 1) before proceeding with his explanation. The trainee suggests that he might have preferred to offer a direct translation to his learner – perhaps deeming this as cognitively less challenging – as he states his inability to use this strategy (*I don't know at all what it is in English*). The trainee proceeds with a verbal explanation and privileges a formal definition (*it's a table which allows you to summarise information*). This is accompanied by a metaphoric gesture (Figure 10, plate 2). It is barely visible in the webcam image but the trainee moves both hands, palms facing each other together. It appears that the gesture is for the trainee himself and helps him to forefront the 'summarising' property of an online mind map that he then uses in the verbal mode. Perhaps in response to there being no change in the facial expression of his learner, Victor then decides to send the learner a link to the mind map image that he had prepared for the activity. The change in which verbal modality is foregrounded is announced (*I'll send you a mindmap*) perhaps to explain why the trainee's gaze has moved from being directed towards the webcam downwards towards the right-hand corner of his screen. This does not engender a change in gaze on behalf of the learner. Potentially this is what prompts the trainee to explicitly announce the modality used (*in the text chat*, Figure 10 plate 5). As he does so, the learner's gaze shifts and he then changes posture away from his webcam whilst we see a movement of his right arm, seemingly towards his mouse.

<p>Victor 25:30- 25:45</p> <p>1</p>	  	<p>eah: je sais pas du tout c' que c'est en anglais\ une carte mentale/ c'est une façon/ c'est un schéma</p> <p>un tableau/ qui te permet de: hum: résumer des informations je vais t'envoyer la carte mentale/ comme ça peut-être que tu pourras euh voir/ de quoi ça a l'air</p>	<p><i>uh: I don't know at all what it is in English\ a mind map/ it's a way/ it's a figure</i></p> <p><i>a table/ that allows you to: hum: summarize information I'll send you the mind map/ that way maybe you'll be able uh to see/ what it looks like</i></p>
<p>Victor text chat 25:46</p> <p>4</p>		<p>http://s12.postimg.org/6gufejrt9/carte_mentale_modifi_e.jpg</p>	<p>http://s12.postimg.org/6gufejrt9/carte_mentale_modifi_e.jpg</p>
<p>Victor 25:46- 25:47</p> <p>5</p>		<p>dans le tchat</p>	<p><i>in the text chat</i></p>

Figure 10: Text chat to send images

Strategies used in the closing phases

Whilst the audio modality was mainly used by the learners to acknowledge the trainees' explanations during the closing phases of learner-initiated sequences, the data analysis shows that these were accompanied visually by both sagittal head movements and facial expressions (Figure 11). However, the data also exemplifies a visual strategy used by the learners to mark the end of the lexical explanation side sequences before continuing with the main task. Learners move proxemically away from the webcam either during or just following the verbal acknowledgement, as if to mark the end of the side sequence within the interaction concerning the task.

Alannah 09:10- 09:11		[ah oh] oui	[ah oh] yes
Sean 25:46- 25:47		ah oui	ah yes

Figure 11: Acknowledgement strategies

The closing sequences were also characterised by the trainees asking one of the learners who had acknowledged the explanation to (re)explain to his/her peer. However, this did not always allow the trainee to check the learner had understood. The data shows instances of the learner, although remaining in the visual mode, moving out of the interactional order and using the face-to-face classroom environment to complete this trainee request (Figure 12).




Adèle 11:30- 11:31 1		tu expliqueras à Catriona/	will <i>you</i> explain to Catriona/
Silence 11:31- 11:38 2			
Inaudible 11:38- 11:44 3		xxx	.xxx

Figure 12: Peer explanations in a different interactional order

In other instances, the learners were asked the translation of the item or for some information about the use of an equivalent term in English, for example whether the formal/informal second-person singular pronoun existed. The requests were accompanied in the visual mode by pensive gaze directed outside of the webcam frame.

Conclusion and Perspectives

Prior work has documented lexical explanations from a conversation analysis approach in face-to-face settings and has described prototypical teaching sequences for both teacher and student-initiated explanations by a ternary model. However, the multimodal reality of the interaction situation is lost in this approach and nuances in the learning process may be overlooked. Whilst multimodal studies into synchronous environments other than written text environments are rare, potentially due to the demands of data capture as well as multimodal transcription, a growing body of research examines webconferencing-based environments from this perspective. Although some research has explored lexical explanations in this context, the topic has been focused upon within the wider context of instructional regulations or peer scaffolding.

This qualitative study concentrated on three trainee teachers' lexical explanations during two sessions of a telecollaborative exchange and examined the different strategies employed during student- and teacher-initiated lexical explanations across the visual mode and text chat and audio modalities. The synchronous interactions provide access to emerging pedagogical and multimodal communication strategies for 'doing vocabulary teaching'. The data illustrated that in the openings of such sequences, the visual mode was used to project active listening strategies during student-initiated sequences whilst in teacher-initiated sequences, the audio modality was used to highlight target items in the TCU final position and the text chat provided a complementary role to help establish common ground concerning the target item. Shifts from the audio to text chat were signalled through gaze shifts which aided conversation management and allowed the trainees to either maintain an ongoing turn or signal the end of a turn.

Analysis of the visual mode in combination with the audio modality shed light on a 'thinking break' between the opening and explanation components of the sequence that trainees used to maintain a sense of coherence. There appears to be a continuum concerning the explicitness of this strategy that ranges from employing a pensive facial expression with gaze off screen and a verbal hesitator to the trainee momentarily removing him/herself from the interactional order. The synchronous nature of the environment requires trainees to offer unplanned, spontaneous explanations and this strategy, which introduces a pause in the interaction, may be one means to avoid cognitive overload.

Trainees employed a range of verbal strategies in their lexical explanations. These were accompanied, firstly, by deictic and iconic gestures that, if made visible in the webcam view for the learners, had a complementary role in their explanations and, secondly, by metaphoric gestures that helped foreground different properties of the lexical item in order to include them in a verbal explanation. The text chat modality was also employed to send hyperlinks to images but again, if used, this change in modality in the verbal mode was always signalled to learners in the visual mode and, if visual cues were not noticed by learners, in the audio modality. In addition, the data showed different strategies employed in the closing phases of the sequence: firstly, the combination of the audio modality and visual mode by learners to acknowledge the explanation offered; and, secondly, a change in learner proximity to the webcam that appears to signal the end of the side-sequence focusing on form and, by indicating a wish to return to the main theme, to offer the trainees indications of how to manage the continuation of the interaction.

The analysis conducted sheds light on trainee teachers' abilities to exploit different communication modes for the specific pedagogical purpose of lexical explanations. Although it provides descriptions of the combination of different communication modes in synchronous interaction for lexical explanations, it does not cover how the multimodal activities are received by the other participants as the interaction unfolds. Indeed a limitation of this study is that it does not include stimulated recall data. Due to the ecological nature of the corpus, where specific research questions stemmed from initial viewings of the data and did not govern its collection, combined with the demands of multimodal transcription, the researcher was confronted with the issue of no longer having access to the participants once the research question had been identified. One way to overcome this would be to employ eye-tracking data collection techniques during the initial synchronous data collection. This would advance our understanding of how participants 'read' the different multimodal interaction and whether modes are read simultaneously or different channels are foregrounded in reception (O'Rourke *et al.*, 2015; Stickler & Shi, 2015). Given that each group in the telecollaborative exchange followed the same lesson plan, additional research on the same corpus has been conducted by other research team members to isolate a collection of lexical explanation sequences that focus on the same target item in order to conduct a comparative analysis of how different trainees manage the same teaching moment (Holt & Tellier, 2017).

As Internet access by different learner populations improves, Higher Education institutions are increasingly interested in offering more flexible teaching and learning delivery methods that are often independent of place. Webconferencing-supported language teaching opens up possibilities of bringing together native speakers and learners of the target language and foreground synchronous interaction skills in distance learning. As the mediation of this language learning is through ever more sophisticated platforms, training online language teachers to become aware of the range of semiotic resources available to support different pedagogical practices, including the provision of lexical explanations, appears increasingly important.

References.

- Altorfer, A., Jossen, S., Würmle, O., Käsermann, M., Foppa, K. & Zimmermann, H. (2000). Measurement and meaning of head movements in everyday face-to-face communicative interaction. *Behavior Research Methods, Instruments & Computers*, 32(1). pp.17-32.
- Bavelas, J. & Chovil, N. (1997). Faces in dialogue. In Russell, J.A. & Fernandez-Dols, J.M. (Eds.), *The psychology of facial expression*. Cambridge: Cambridge University Press. pp.334-347.
- Bavelas, J., Gerwing, J. & Healing, S. (2014). Hand and Facial Gestures in Conversational Interaction. *The Oxford Handbook of Language and Social Psychology*. Oxford: Oxford University Press. pp.111-130.
- Bétrancourt, M., Guichon, N. and Prié, Y. (2011). Assessing the use of a Trace-Based Synchronous Tool for distant language tutoring. *Proceedings of the 9th International Conference on Computer-Supported Collaborative Learning*, Hong-Kong, July 2011. pp.478-485.
- Cappellini, M. (2013). Co-construction des routines d'étayage dans un tandem franco-chinois par visioconférence. In Dejean-Thircuir, C., Mangenot, F., Nissen, E., & Soubrié, T. (Eds). *Actes du colloque EPAL 2013*.

- Chanier, T. & Vetter, A. (2006). Multimodalité et expression en langue étrangère dans une plateforme audio-synchrone. *Apprentissage des Langues et Systèmes d'Information et de Communication (ALSIC)*, 9. pp.61-101.
- Codreanu, T. & Combe Celik, C. (2013). Effects of webcams on multimodal interactive learning. *ReCALL*, 25(1). pp.4-29.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Develotte, C., Guichon, N. and Vincent, C. (2010). The use of the webcam for teaching a foreign language in a desktop videoconferencing environment. *ReCALL*, 23 (3). pp.293-312.
- Fasel Lauzon, V. (2008). Interactions et apprentissages dans des séquences d'explication de vocabulaire. *Travaux neuchâtelois de linguistique*, 48. pp.83-104.
- Goodwin, M.H. & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, 62(1/2). pp.51-75.
- Guichon, N., Bétrancourt, M. and Prié, Y. (2012). Managing written and oral negative feedback in a synchronous online teaching situation. *Computer Assisted Language Learning*, 25(2). pp.181–197.
- Guichon, N., Blin, F., Wigham, C.R., & Thouëšny, S. (2014) *ISMAEL LEarning and TEaching Corpus*. Dublin, Ireland: Centre for Translation and Textual Studies & Lyon, France: Laboratoire ICAR.
- Guichon, N. & Cohen, C. (2014). The Impact Of The Webcam On An Online L2 Interaction. *Canadian Modern Language Review*, vol. 70, n°3. pp.331–354.
- Guichon, N. & Cohen, C. (2016). Multimodality and CALL. In Farr, F & Murray, L. (Eds.). *The Routledge Handbook of Language Learning and Technology*. London: Routledge. pp.509-521.
- Guichon, N. & Drissi, S. (2008). Tutorat de langue par visioconférence : comment former aux régulations pédagogiques. *Les Cahiers de l'ACEDLE*, Vol. 5, n°1. pp.185-217.
- Guichon, N. & Wigham, C.R. (2016). A semiotic perspective on webconferencing-supported language teaching. *ReCALL Journal*, 28(1). pp.62-82.
- Gulich, E. (1990). Pour une ethnométhodologie linguistique. Description des séquences conversationnelles explicatives. In Charolles, M, Fisher, S. & J. Jayez (Eds.). *Le discours. Représentations et interprétations*. Nancy: Presses Universitaires de Nancy. pp.71-109.
- Halliday, M.A.K. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.
- Hampel, R. (2003). Theoretical perspectives and new practices in audio-graphic conferencing for language learning. *ReCALL*, 15(1). pp.21–36.
- Hampel, R. and Stickler, U. (2012). The use of videoconferencing to support multimodal interaction in an online language classroom. *ReCALL*, 24(2). pp.116–137.
- Hauck, M. & Youngs, B. L. (2008). Telecollaboration in multimodal environments: the impact on task design and learner interaction. *Computer Assisted Language Learning*, 21(2). pp.87-124.
- Holt, B. & Tellier, M. (2017) Conduire des explications lexicales. In Guichon, N. & Tellier, M. (Eds). *Enseigner l'oral en ligne : Une approche multimodale*. Paris: Didier. pp.59-90.
- Kendon, A. (1982). The study of gesture: some observations on its history. *Recherches Semiotique/Semiotic Inquiry*, 2(1). pp.25-62.
- Knight, D. (2011). *Multimodality and Active Listenership*. London: Continuum.
- Kress, G. (1993). Against arbitrariness: the social production of sign as a foundational issue in critical discourse analysis. *Discourse and Society*, 4(2). pp.169-191.
- Kress, G. (2003). *Literacy in the new media age*. London: Routledge.
- Kress, G. (2010). *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London. New York: Routledge.
- Kress, G. & Van Leeuwen, T. (1996) *Reading Images: The Grammar of Visual Design*. Routledge: New York.

- Lamy, M.-N. (2012). Click if You Want to Speak: Reframing CA for Research into Multimodal Conversations. *Online Learning, International Journal of Virtual and Personal Learning Environments*, 3(1). pp.1-18.
- Lamy, M-N. (2012b). Diversity in modalities. In Stockwell, G. (Ed.) (2012). *Computer-Assisted Language Learning: Diversity in Research and Practice*. Cambridge: Cambridge University Press. pp.110-127.
- Lazaraton, A. (2004). Gesture and speech in the vocabulary explanations of one ESL teacher. A microanalytic inquiry. *Language Learning*, 54(1). pp.79-117.
- Faraco, M. & Kida, T. (2008). Gesture and the Negotiation of Meaning in a Second Language Classroom. In McCafferty, S. & Stam, G. (Eds.) *Gesture. Second Language Acquisition and Classroom Research*. New York: Routledge. pp.280-297.
- Levy, M., & Stockwell, G. (2006). *CALL dimensions. Options and issues in Computer-Assisted Language learning*. Mahwah, NJ: Lawrence Erlbaum Associates.
- McNeill, D. (1992). *Hand and Mind: What gestures reveal about thought*. Chicago: The University of Chicago Press.
- Mortensen, K. (2011). Doing Word Explanation in Interactio. In Pallotti, G. & Wagner, J. (Eds). *L2 learning as social practice: Conversation-analytic perspectives*. Honolulu, HI: University of Hawaiï, National Foreign Language Resource Center. pp.135-162.
- O'Rourke, B., Prendergast, C., Shi, L., Smith, B. & Stickler, U. (2015). Eyetracking in CALL: present and future. In Gimeno Sanz, A.-M. , Levy, M., Blin, F. & Barr, D. (Eds.), *WorldCALL: Sustainability and Computer-Assisted Language Learning*. London: Bloomsbury. pp. 285-298.
- Sacks, H. Schegloff, E.A. & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*. 50(4). pp.696-735
- Satar, M. (2013). Multimodal language learner interactions via desktop videoconferencing with a framework of social presence: Gaze. *ReCALL*, 25(1). pp.122-142.
- Schegloff, E.A., Jefferson, G. a Sacks, H. (1977). The Preference for Self-Correction in the Organization of Repair in Conversation. *Language*, 53(2). pp. 361-382.
- Sloetjes, H. & Wittenburg, P. (2008). Annotation by category – ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.
- Stickler, U. & Shi, L. (2015). Eye movements of online Chinese learners. *CALICO*, 32(1). pp. 1-30.
- Tellier, M. & Stam, G. (2010). Découvrir le pouvoir de ses mains : La gestuelle des futurs enseignants de langue. *Faculty Publications*. Paper 33.
- Tsuchiya, K. (2013). *Listenership Behaviours in Intercultural Encounters*. John Benjamins Publishing Company: Amsterdam.
- Wang, Y. (2006). Negotiation of meaning in desktop videoconferencing-supported distance language learning. *ReCALL*, 18(1). pp.122-145.
- Wang, Y. (2007). Task Design in Videoconferencing-Supported Distance Language Learning. *CALICO*, 24(3). pp.591-630.

Appendices

Appendix A – Audio modality transcription conventions

learner turn
trainee-teacher turn

bold verbal utterance during which the screen capture shot was taken

/ rising intonation

\ falling intonation

+ one-second pause

++ two-second pause

+++ three-second pause

: lengthening of the last sound

“ ” code-switching

[point at which overlap by another speaker starts

] point at which overlap by another speaker stops

() extralinguistic features e.g. (cough)

xxx inaudible

Acknowledgements

The author is grateful to the ASLAN project (ANR-10-LABX-0081) of Université de Lyon, for its financial support within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) of the French government operated by the National Research Agency (ANR). This research was also supported by the Ulysses programme funded jointly by the Irish Research Council and the French Ministry of Foreign Affairs. Thanks to reviewers for their comments on the initial version of this paper.