ReCALL (2022), 34: 3, 255–273
doi:10.1017/S0958344022000076

RESEARCH ARTICLE

Multimodality in webconference-based language tutoring: An ecological approach integrating eye tracking

Marco Cappellini
Aix-Marseille University, Laboratoire Parole & Langage UMR 7309 CNRS, France (marco.cappellini@univ-amu.fr)

Yu-Yin Hsu
The Hong Kong Polytechnic University, Hong Kong, China (yyhsu@polyu.edu.hk)

Abstract
Drawing on existing research with a holistic stance toward multimodal meaning-making, this paper takes an analytic approach to integrating eye-tracking data to study the perception and use of multimodality by teachers and learners. To illustrate this approach, we analyse two webconference tutoring sessions from a telecollaborative project involving pre-service teachers and learners of Mandarin Chinese. The tutoring sessions were recorded and transcribed multimodally, and our analysis of two types of conversational side sequences shows that the integration of eye-tracking data into an ecological approach provides richer results. Specifically, our proposed approach provided a window on the participants' cognitive management of graphic and visual affordances during interaction and uncovered episodes of joint attention.

Keywords: webconferencing; online tutoring; telecollaboration; eye tracking; affordances; multimodality

1. Introduction

Webconferencing has emerged as an important tool in computer-assisted language learning (CALL) programs (Hubbard, 2017) and especially in telecollaboration (O'Dowd, 2018). By *webconferencing*, we refer to tools that allow visual and audio communication over the internet, usually alongside other forms of communication and collaboration, such as text-based chatrooms and/or screen sharing. Common platforms include client software such as Skype, or online platforms such as Adobe Connect and, more recently, Zoom.

Webconferencing has been studied in terms of the modalities it integrates, beginning with seminal work by Develotte and colleagues on the use of images and individual internet users' webcams (Develotte, Guichon & Vincent, 2010), and on conversational phenomena such as multimodal conversational openings gestures and proxemics (Develotte, Kern & Lamy, 2011). The presence of multiple modalities, in videoconferencing as well as in other computer-mediated communication (CMC) environments, has been linked to teachers' and learners' communicative capacity to draw on them, and this has been conceptualised in various ways, including (multi) literacies (Fuchs, Hauck & Müller-Hartmann, 2012) or multimodal competence (Hauck, 2010).

The present article relates recent research in this area to a new methodological framework for studying language teacher education and specifically the development of techno-pedagogical competence in online tutoring (Guichon & Cohen, 2016). As such, it is concerned with several important but understudied aspects of CALL that were identified in a recent review by Gillespie

Cite this article: Cappellini, M. & Hsu, Y.-Y. (2022). Multimodality in webconference-based language tutoring: An ecological approach integrating eye tracking. *ReCALL* 34(3): 255–273. https://doi.org/10.1017/S0958344022000076

© The Author(s), 2022. Published by Cambridge University Press on behalf of European Association for Computer Assisted Language Learning.

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press

(2020), such as multimedia, online learning, and content and language integrated learning. We demonstrate this methodological approach in a case study of a telecollaboration between learners and pre-service teachers of Mandarin Chinese as a foreign language, with the latter in the role of online language tutors (Cappellini & Hsu, 2020). Our aim is to support an ecological approach by shedding new light on the perception and the use of multimodality by future teachers with the help of eye-tracking data (Stickler, Smith & Shi, 2016). In section 2, we review the relevant recent literature linking this approach to webconferencing. In section 3, we provide information on the context, the participants and the methods used in our case study. In section 4, we present and discuss our results before concluding.

2. Literature review

2.1 Multimodality

Following a long tradition in CALL (see, for instance, Lamy & Hampel, 2007), we define *multimodality* as the simultaneous presence of multiple modes of communication. *Modes* are defined as semiotic resources or semiotic regimes that interlocutors can use to co-construct meaning (Bezemer & Kress, 2016). In webconferencing, such modes include, but are not limited to, written language, gestures, facial expressions, proxemics and oral language in its verbal aspects as well as prosody, pitch and delivery (Rivens Mompean & Cappellini, 2015). These modes can be present at any time during the use of webconferencing, but interlocutors do not necessarily perceive them as relevant to expressing their intentions. We will draw on the concept of affordance (Blin, 2016; Gibson, 1979), which refers to the perception of an actor engaged in an action of the enabling and/or constraining effects of elements of the environment on that action (see below).

Studies on the multimodality of webconferencing interactions in CALL have utilised one of two broad approaches: one analytic and the other one holistic. In analytic approaches, researchers focus on one mode and study it in isolation from the others. For instance, Yamada and Akahori (2010) manipulated the presence of the video feed of one of the interlocutors to gauge its impact on the other interlocutors' grammatical accuracy correction while speaking. Although this approach has informed some recent studies (Kozar, 2016, for instance), research has more commonly adopted holistic approaches in which multimodality is conceived of as a whole and is studied within paradigms such as multimodal conversation analysis (CA) (Cappellini & Azaoui, 2017; Sert & Balaman, 2018), interactional sociolinguistics (Satar, 2016), social semiotics, or combinations thereof (Helm & Dooly, 2017; Satar & Wigham, 2017). These studies have enhanced our understanding of how multimodality is used as a whole during interaction, often by focusing on particular conversational dynamics such as instruction-giving sequences (Satar & Wigham, 2017, 2020), policing (Sert & Balaman, 2018), or side sequences of negative feedback (Cappellini & Azaoui, 2017). This article aims at proposing a methodological approach that starts with multimodal CA and articulates it with an ecological approach based on the concept of affordance and integrating relevance theory. In broadening multimodal CA in this way, we aim to gain insights into the cognitive dimensions of interaction. We take techno-pedagogic competence as a test case for this methodological approach.

2.2 Techno-pedagogic competence

Previous studies on multimodality have often incorporated reflections on the competencies learners and teachers must develop if they are to take part in online interactions effectively. Several models of pedagogical competence to teach languages with ICTs have been proposed (Dooly, 2010; Kessler, 2016, among others). One of the most influential frameworks is Hampel and Stickler's pyramid (2005, 2015). Although its originators studied webconferencing (Hampel & Stickler, 2012), the pyramid framework was not specifically conceived for this type

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press

of CMC. On the other hand, Guichon's (2012) framework of techno-semio-pedagogical competence or techno-pedagogic competence (Guichon & Cohen, 2016) was designed from the outset to describe the integration of ICTs into language teaching, and has subsequently been extensively adapted to teaching through webconferencing (Guichon & Tellier, 2017). Guichon defines techno-semio-pedagogical competence as "knowledge and skills concerning:

- the communication tools available (forums, wikis, videoconferences, etc.) that are best suited to the objectives of a given teaching sequence;
- the appropriate choice of modes (written, oral, video, or a combination) for a given activity and for the development of linguistic competences;
- the pedagogical management of learning activities where CMC tools are central or incidental (planning, regulating task implementation, evaluating learning)" (Guichon, 2012: 187; our translation).

Importantly, however, there is a gap between the definitions of techno-pedagogical competence and the methodological tools to study it. In fact, most authors agree that such competence includes not only the ability to effectively use relevant modes and strategies for communication (and perhaps teaching; Dooly, 2016) but also knowledge and awareness of semiotic modes (Guichon, 2012; Hauck, 2010). Such knowledge/awareness has often been studied through retrospective introspection, especially in the form of learning logs (e.g. Fuchs *et al*., 2012), and less often through stimulated recall (Cohen, 2017), but never within (inter)action itself. Recent advancements in eye-tracking data collection and related methodologies have made it possible to fill this gap in methodology.

2.3 Ecological approaches and affordances

Following Cappellini (2021), we developed an ecological approach based on the work of Bronfenbrenner (1979) and van Lier (2004). Bronfenbrenner defines ecological approaches in opposition to experimental methodologies. He requires that participants be able to manipulate the environment (ecological validity) and that their perspectives be included in research (phenomenological validity). Following van Lier, we conceive of the environment as a multimodal reservoir of semiotic resources that are drawn upon to make meaning (Bezemer & Kress, 2016). In this study, we focus on the relationship between the participants and their environment on the screen in terms of perception of the elements of the environment and their use in interaction. Participants' perception of the *elements* in the environment is dependent on the actions they carry out, as well as their interpretations. We conceive of these as affordances in the sense of Blin's (2016) post-cognitive approach. In other words, affordances are not given before the action or interaction but emerge during the interaction with/through an environment, including digital environments. Thus we draw on ecological approaches to focus on the dialogic relationship between agents and their environment based on the concept of affordance.

2.4 Relevance theory

Relevance theory was first proposed by Sperber and Wilson (1986) as a framework for studying the cognitive dimension of human interaction. The framework allows the investigation of the interplay between communicative behaviour and contextual factors to explain how interlocutors' attention is managed. In this framework, *context* can be roughly defined as what is mutually manifest to interlocutors. Elements of the environment can be rendered mutually manifest through *ostensions* – that is, communicative behaviour. Drawing on the interlocutor's ostensions and on representations of their intentions, one we can process information in order to formulate an interpretation of what the interlocutor is doing and/or is trying to communicate. In this study,

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press

we conceive ostensions as drawing on communicative affordances that emerge during interaction and allow interlocutors to construct interpretations about each other's actions and intentions. These ostensions therefore draw on the multimodality of the webconferencing environment. The perception of such an environment is at the core of this study.

2.5 An ecological approach to multimodality integrating eye tracking

In a previous study (Cappellini & Hsu, 2018), we argued that the emergence of affordances could be studied in part using eye-tracking data. Eye-tracking techniques have been used to collect data in CALL and telecollaboration for about a decade (O'Rourke, 2012), but only recently have technological advancements allowed researchers to deploy it to study webconferencing environments (Stickler *et al.*, 2016), answering various calls to do so from different researchers (Guichon & Wigham, 2016; Sert & Balaman, 2018). Specifically, eye-tracking technology can collect data about a subject's gaze fixation (in our case, on a screen), allowing us to investigate where the subject is looking at each moment of an interaction. The eye–mind hypothesis (Conklin, Pellicer-Sánchez & Carrol, 2018) holds that gaze fixations can provide insights into a subject's cognitive processes. Accordingly, we formulate the hypothesis that if tutors establish a relationship with the environment on the screen in cycles of action-perception-interpretation (van Lier, 2004), then studying how they scrutinise the multidimensional semiotic space of the screen can provide information about their *knowledge and awareness* of the elements of the digital interfaces they use, which is a key component of their techno-pedagogic competence. Within this theoretical framework, we aim to answer the following main research question: what are the contributions and limits of eye tracking for studying the perception and use of multimodality during webconferencing? A secondary research question is, what multimodal conversational dynamics are apparent in conversation side sequences for scaffolding (Cappellini, 2016)?

3. Context and methods

3.1 Context and participants

Our data were collected during a telecollaboration between L1 French learners and pre-service teachers of Mandarin Chinese as a foreign language. The teachers were first-year MA students at the Hong Kong Polytechnic University. They came from Mainland China and also speak English (perhaps among other languages). The learners were second-year bachelor's students in Chinese Language, Literature and Civilisation at Aix-Marseille University, with proficiency ranging between A2 and B1 (Council of Europe, 2001). The telecollaboration took place in three iterations, during the spring semesters of 2017, 2018 and 2019, and involved 18 teacher–learner groups. Each year, the future teachers received instructions about the French students' Chinese language curriculum and developed conversational activities to be carried out during their webconferencing sessions. The interested reader will find more information about the pedagogical set-up in Cappellini and Hsu (2020).

3.2 Data collection and the corpus of the study

The current study used two randomly chosen webconferencing sessions as a test case for our methodological approach. The first session was drawn from the first iteration of the telecollaboration, which involved one pre-service teacher and two learners of Chinese. Data were collected from the teacher's side of the exchange only. The second session was part of the second iteration, which involved one pre-service teacher and a single learner. In this case, data were collected from both sides. All the learners in the two sessions were at A2 level in Mandarin Chinese. Written consent for the study was obtained before each participant joined the recorded webconference sessions.

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press

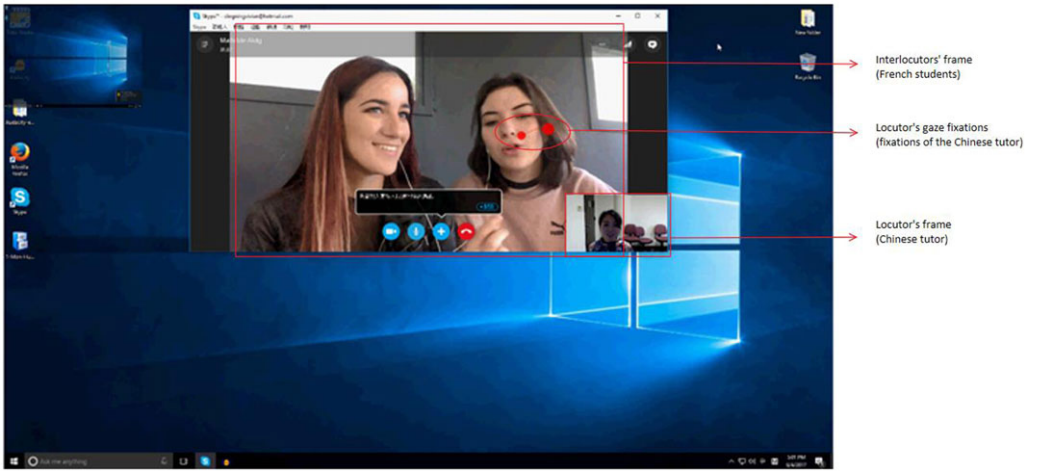


Figure 1. Example of a screen capture with eye-tracking data

We used a Tobii X300 eye tracker with Tobii Pro Studio software to collect the data. To ensure the webconferencing setting in our study was similar to typical online language-tutoring sessions, we asked the participants to sit comfortably at a distance of about 65 cm from the recording screen and face it directly. Then we conducted a nine-point calibration of the eye tracker for gaze direction. Unlike the laboratory experimental eye-tracking studies, participants were not restricted in terms of head movement. There were three parts to each dataset:

1. The single-channel audio stream comprising audio from all the participants;
2. A dynamic screen capture of participants' eye movements during the session, for the tutor only in the first case and for both parties in the second case; and
3. An eye-tracking recording set at 120 Hz and thus providing 120 captures of each participant's gaze position every second.

The first session recording lasted 54 minutes (1,028 turns, 4,241 words) and the second, 34 minutes (567 turns, 3,868 words).

Data were then exported in formats compatible with the EUDICO Linguistic Annotator (ELAN; Sloetjes & Wittenburg, 2008). Eye-tracking data were exported in two ways. First, gaze fixation and gaze path were presented as dots and lines on the dynamic screen capture. Figure 1 shows an example screen capture from the first session. Second, we defined three areas of interest: the written chat, the participants' own camera feed, and the face(s) in the camera feed of their interlocutors. Fixations on these areas of interest were then exported into ELAN for annotations, with a separate annotation tier assigned to each area of interest.

The audio recordings were transcribed by the authors collaboratively in ELAN, with a tier for each participant. For the verbal material, we used the transcription convention adopted in Cappellini (2021), an adaptation from the ICOR convention, a standard for interaction research in France.[1]

3.3 Data analysis

The audio stream and the video recording(s) with gaze dots were integrated into the ELAN interface. Other than for transcription of verbal data, we used an adaptation of Wigham

[1]https://www.itereva.org/mompepe/frontend/images/2013_Conv_ICOR.pdf

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press

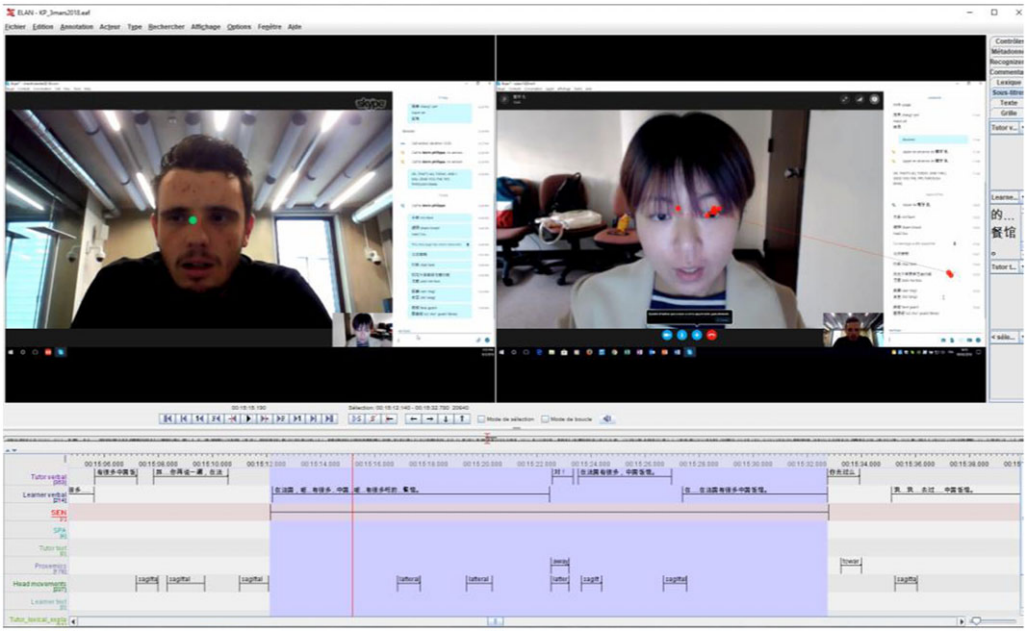


Figure 2. Example of annotations in ELAN

(2017) already presented in Cappellini and Hsu (2018) to manually annotate multimodal elements, including gestures, proxemics, head movements, participation in the chat window, and actions.

Our main focus was on two types of conversational side sequences that have previously been found relevant in webconferencing settings for language learning (Cappellini, 2016). The first are sequences of potential acquisition (hereafter SPA; de Pietro, Matthey & Py, 1989), which correspond to conversational side sequences (Jefferson, 1972) in which the learner faces a gap in their competence, usually a missing lexical item, and solicits help from their interlocutor. The second are sequences of normative evaluation (hereafter SEN; Py, 2000): another type of side sequence in which the language expert considers that there has been an error in the interlocutor's expression and signals this.[2]

Analysis was conducted by replicating the procedure in previous research (Cappellini & Azaoui, 2017). First, the two authors independently identified the conversational phenomena of interest, then discussed any discrepancies until agreement was reached. Next, we analysed each instance from a multimodal interaction perspective, informed by "embodied" CA (Goodwin, 2000; Mondada, 2016), including a focus on gaze trajectories to understand management of interaction. Finally, we compared our analyses of the different instances to highlight common patterns. Figure 2 shows an example screenshot of annotations in ELAN taken from our second example.

In the analysis, we investigated how the interlocutors directed their focal attention to the elements of the screen during the unfolding conversations that manifested either type of side sequence while using sets of affordances to co-construct meaning in terms of ostensions and inferences. Our aim was primarily methodological, in that we wanted to assess the contributions and the limits of the approach we propose. Inevitably, however, we gained some insights into the

[2]The conversational side sequences we adopt as units of analysis are defined in relation to the Francophone interactionist literature (see, for instance, Pekarek Doehler, 2000), which is based on a combination of conversation analysis and the Vygotskian framework. The Anglophone reader can find a thorough discussion of this, under the label of the *strong socio-interactionist perspective*, in Mondada and Pekarek Doehler (2004).

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press

Table 1. Overview of the side sequences

	Sequences of potential acquisition	Sequences of normative evaluation	Total
Session 1	16	10	26
Session 2	4	7	11
Total	20	17	37

cognitive process at work while interlocutors, particularly the pre-service teachers, deployed their techno-pedagogic competence, and these should be explored using larger datasets in the future.

4. Results and discussion

In this section, we first provide an overview of all side sequences of interest in the two webconferencing sessions, and then we focus on one example for each type to offer a more comprehensive analysis.

4.1 Overview

As Table 1 shows, we identified 37 side sequences in the corpus, 20 SPA, and 17 SEN.

The three interlocutors in the first session generated more side sequences than the pair in the second session, with SPA especially frequent at 62% of all Session 1 side sequences, and 80% of SPA overall. This cannot be explained entirely by the fact that this session lasted longer and produced more verbal content than the other; rather, it was at least partly related to other differences. The tutor in Session 1 had a more conversational approach, asking the learners questions that introduced crossed expertise, where the learners were topic experts and the tutor the language expert, an approach previously found to generate more SPAs (Cappellini, 2016). The tutor in Session 2, on the other hand, adopted a more teacher-like posture, producing mainly the initiation-response-feedback (IRF) exchanges that are typical of classroom interaction (Sinclair & Coulthard, 1975), which are likely to lead to SEN in the case of form or content problems.

4.2 Sequences of potential acquisition

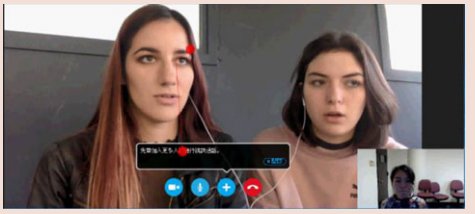
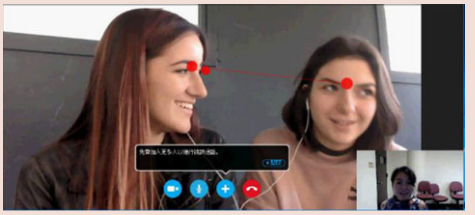
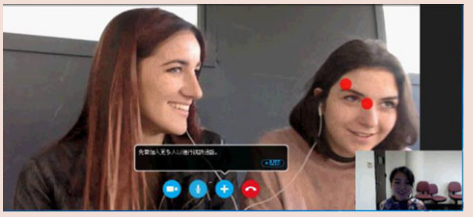
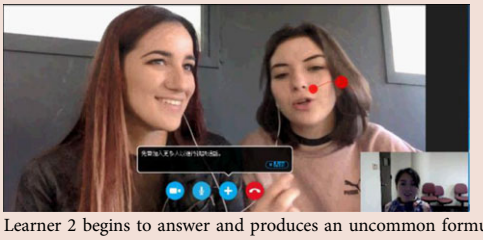
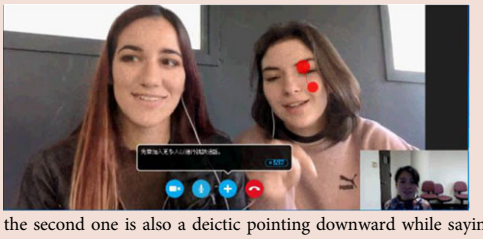
In this section, we present our analysis of a representative excerpt of a Session 1 SPA focusing on multimodal conversation strategies, both by the tutor in terms of techno-pedagogic competence and by the learners in terms of multimodal competence. Table 2 gives a multimodal transcription with our description.[3] The excerpt included both Mandarin Chinese and French, which we translated in parentheses. This side sequence took place after the tutor asked the learners about their location. Learner 1 is on the left.

Although longer than other SPAs in the dataset, this example is representative of the multimodal strategies at work in terms of several characteristics. First, the video feeds emerge as an affordance used by the tutor to understand the learners' orientation and engagement in the interaction. When an interlocutor takes the floor, the tutor looks at the speaker, including when she herself is speaking (though, as we shall see, this dynamic was not observed in Session 2). During TRPs, the tutor looks at the learners to see if they intend to take the floor. For instance, before turn 10, the learners' ostension of looking downwards leads the tutor to the inference that they are not willing to take the floor at that point. Moreover, in case of overlap, the tutor systematically leaves the floor to the learners. The video's status as an affordance whereby the tutor can scrutinise

[3]We strongly encourage the reader to watch the video of this excerpt at https://amupod.univ-amu.fr/video/2432-telecollaboration-polyu-amu/

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press

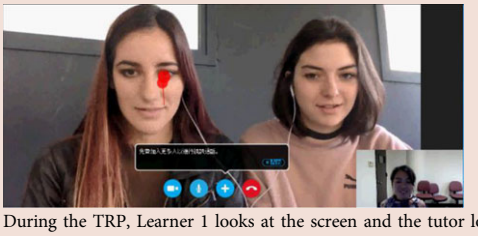
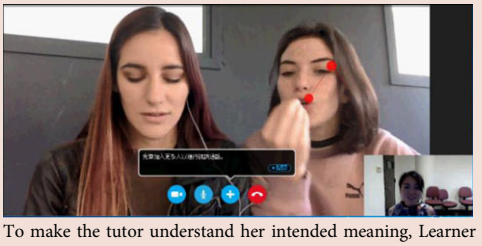
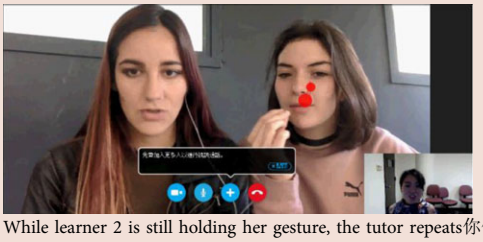
Table 2. Annotated multimodal transcription of a side sequence from Session 1

Turn	Participant and time code	Screenshots	Speech (translation)
1	Learner 1 verbal 00:01:52.526 – 00:01:54.851	 The two learners have a "doing thinking" facial expression while searching for a way to express themselves.  Learner 1 attempts an utterance but stops it after the verb and turns to her fellow learner and whispers (inaudible in the audio recording). Meanwhile, the tutor's gaze is at first on Learner 1's face, and then she turns toward Learner 2, following whoever takes the floor in the conversation between the two learners.	是xxx (it's xxx)
2	Tutor verbal 00:01:54.990 – 00:01:55.280	 The tutor also starts an utterance at turn 2, but she leaves the floor when she sees that Learner 2 is about to speak.	你 (you)
3	Learner 2 verbal 00:01:55.380 – 00:01:56.460	 Learner 2 begins to answer and produces an uncommon formulation in Mandarin Chinese while producing two gestures: the first one is a deictic toward herself while saying 我们 (we);  the second one is also a deictic pointing downward while saying 那里 (where) to express "here".	我们吃那里 (there where we eat)

(*Continued*)

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press

Table 2. (*Continued*)

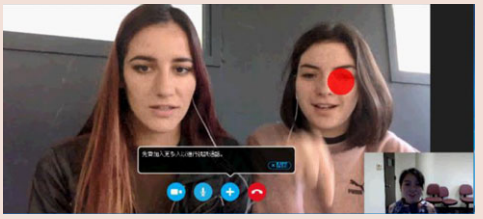
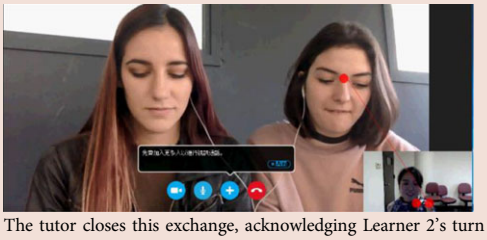
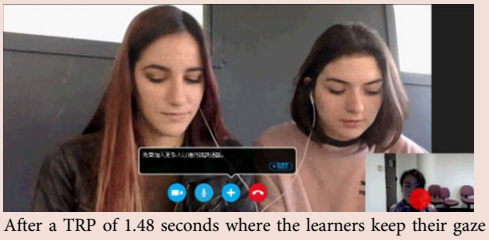
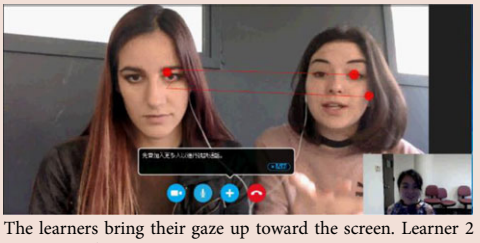
Turn	Participant and time code	Screenshots	Speech (translation)
4	Tutor verbal 00:01:56.870 – 00:01:58.780	 The tutor reformulates Learner 2's turn 3 utterance while nodding. In doing so, she changes the meaning, which signals a misunderstanding, but provides a correct form of the utterance. While the tutor speaks, her gaze follows whichever learner takes the floor and she also looks at her own camera feed, before going back to the interlocutors' faces at the transition relevance place (TRP).[4]	嗯你们在那里 (em you are there)
5	Tutor verbal 00:01:59.940 – 00:02:00.640	 During the TRP, Learner 1 looks at the screen and the tutor looks at her, while Learner 2 changes her gaze from looking downward to looking up at the screen, which attracts the tutor's gaze. The silence after turn 4 lasts for 1.14 seconds, an unusual length, which, coupled with learners' gaze, suggests to the tutor that they did not understand and results in the tutor taking the floor again to start repeating the same utterance. Turns 5 and 6 overlap, which results in the tutor leaving the floor to the learners once more.	你们在- (you are-)
6	Learner 2 verbal 00:02:00.280 – 00:02:01.075	 To make the tutor understand her intended meaning, Learner 2 takes the floor to repeat 我们吃 (we eat), while producing an emblem as a gesture to reinforce the meaning of "eat". In this case, it is not possible to say whether the tutor's gaze is directed at the emblem or at Learner 2's face, since the two are too close.	我们吃 (we eat)
7	Tutor verbal 00:02:01.475 – 00:02:02.325	 While learner 2 is still holding her gesture, the tutor repeats你们吃- (you eat-), acknowledging that she has understood. In this turn, it is not clear whether the tutor once again leaves the floor without completing her utterance because she sees Learner 1 repeating Learner 2's gesture or whether the utterance is complete.	你们吃- (you eat-)

(*Continued*)

[4]A TRP is a point in conversation, usually a silence, when one or more interlocutors think the previous turn ended and the floor is open to be taken.

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press

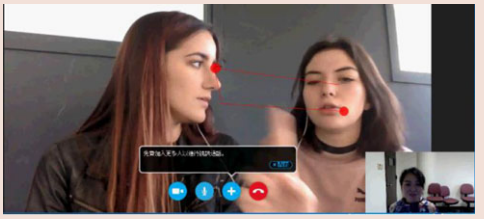
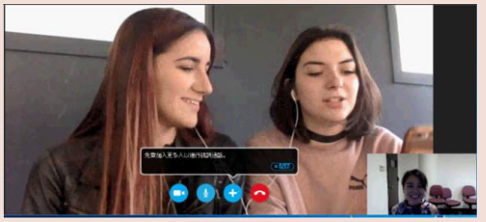
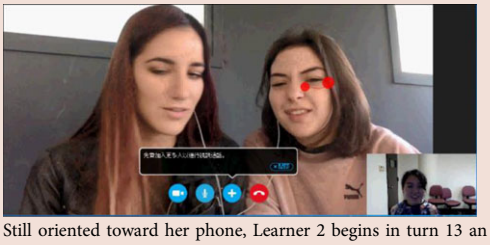
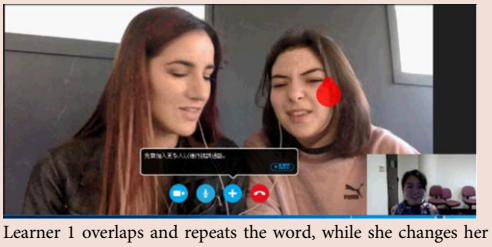
Table 2. (*Continued*)

Turn	Participant and time code	Screenshots	Speech (translation)
8	Learner 2 verbal 00:02:02.840 – 00:02:03.640	 Now that the meaning of "eat" is clear, Learner 2 links it back to the issue of their location, producing a more grammatical form of the utterance in turn 3, accompanied by the deictic gesture for "here".	吃饭那里 (where (we) eat)
9	Tutor verbal 00:02:03.985 – 00:02:05.055	 The tutor closes this exchange, acknowledging Learner 2's turn 8 by repeating it, while looking at her own image.	嗯吃饭那里 (em where (you) eat)
10	Tutor verbal 00:02:06.565 – 00:02:09.935	 After a TRP of 1.48 seconds where the learners keep their gaze mostly downward, the tutor again takes the floor. Turn 10 starts with the paraverbal 哦 (oh), maintained for 0.8 seconds. This is followed by the proposal of a word to express the learners' meaning. The sentence is uttered with a short intra-turn pause between 你们在 (you are at) and 餐厅 (the restaurant), to detach the latter word from the rest and draw the learners' attention to it. Once again, the tutor looks at her own image as she speaks, then moves the gaze toward her interlocutors at the TRP.	哦你们在+餐厅 (oh you are at the restaurant)
11	Learner 2 verbal 00:02:10.845 – 00:02:13.065	 The learners bring their gaze up toward the screen. Learner 2 switches to French to say "yes 餐厅 that's it" (which is not understood by the tutor, as she does not speak French). She accompanies this part of the turn with a deictic gesture toward the screen, probably with the meaning "what you said", which specifies the verbal anaphora *c'est ça* (that's it).	oui餐厅c'est ça + non: non non (yes restaurant that's it + no no no)

(*Continued*)

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press

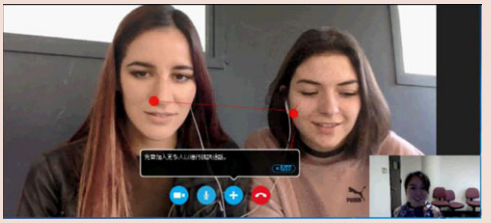
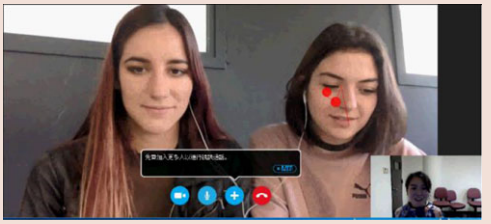
Table 2. (*Continued*)

Turn	Participant and time code	Screenshots	Speech (translation)
		 After a short intra-turn pause while the learners turn toward each other, and while the tutor's gaze rests between the two, Learner 2 repairs what she just said and repeats three times "no" while producing the emblem gesture for "no".	
12	Learner 2 verbal 00:02:14.650 – 00:02:15.010	 During the following silence, Learner 1 keeps her orientation toward Learner 2; Learner 2 nods while producing a facial expression of difficulty and a gesture that can be interpreted as an emblem for "almost" or "not exactly"; the tutor smiles and gives the learners time. While producing a filler (嗯 em), Learner 2 takes her cell phone and probably uses an online dictionary. Both learners are oriented toward the cell phone.	嗯 (em)
13	Learner 2 verbal 00:02:16.175 – 00:02:19.090	 Still oriented toward her phone, Learner 2 begins in turn 13 an utterance to propose an alternative word to describe the place where they are. After an intra-turn pause, she turns toward the screen and leans forward while saying the word 餐馆 (school cafeteria) with a facial expression denotating incertitude.	这种+餐馆+餐馆 (this type + school cafeteria)
14	Learner 1 verbal 00:02:17.970 – 00:02:18.530	 Learner 1 overlaps and repeats the word, while she changes her orientation from the cell phone to the screen.	餐馆 (school cafeteria)

(*Continued*)

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press

Table 2. (*Continued*)

Turn	Participant and time code	Screenshots	Speech (translation)
15	Tutor verbal 00:02:19.185 – 00:02:19.815	 Then the tutor repeats in turn 15 the word 餐馆 and opens a TRP during which both learners nod positively.	餐馆 (school cafeteria)
16	Tutor verbal 00:02:21.105 – 00:02:21.785	 The tutor then repeats the word again, while the learners lean back slightly, which closes the conversation side sequence.	餐馆 (school cafeteria)

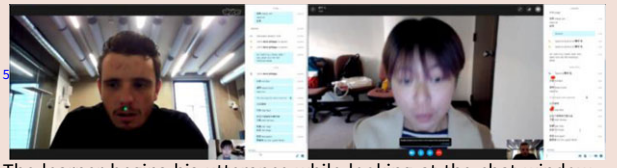
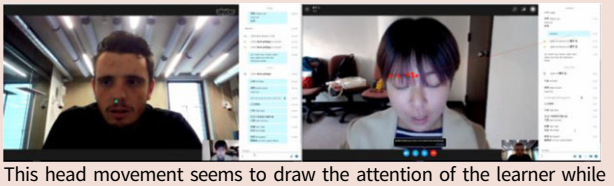
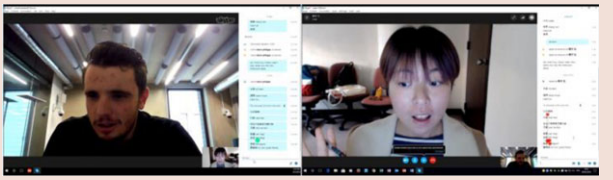
learners' orientation and gaze is especially evident in turns 1–3. In other words, gaze and the interlocutor's gaze perception function as interactional gestures – that is, gestures to manage interaction (Bavelas, Chovil, Coates & Roe, 1995).

Second, it is worth noting that the gestures produced in both sessions are mainly deictics and emblems. The latter are defined as culturally specific gestures whose meaning can be understood without any additional element like speech (Kendon, 1982). The tutor does not direct the focus of her attention toward such gestures, which remain in the periphery of focal attention. The only exception occurred in Session 1, when both learners patted Learner 1's shoulder while soliciting the word "back" in a side sequence about backache. This absence of focal attention on gestures may be linked to the affordances of the webconferencing environment and to more general patterns of focal attention on gestures. Gullberg and Holmqvist (2006) showed that overt focal attention to gestures is present when the gesture is in the extreme peripheral area, very far from the speaker's face. In webconferencing settings, such gestures would be outside the frame of the camera, either invisible, as here, or reduced to more central areas (Holt & Tellier, 2017). In either case, the need for or possibility of focal attention is excluded.

A third characteristic of this type of sequence, evident in both sessions, is the fact that affordances for successful communication are not restricted to software features, but include other technical artefacts in the interlocutors' physical environments, particularly the learners. The main affordance is a cell phone, which is used to access the internet for more information (e.g. a translation, as in the example above). Fourth, the aural mode conveyed by the audio modality emerges as an affordance to highlight parts of the utterance. As seen in turn 10 in Table 2, this occurred through the use of intra-turn pauses that detach a lexical item from the whole utterance. This indicates a multimodal strategy at work, especially in Session 2, where it is also combined with the use of the chat window. Lastly, there is a difference between Session 1, where the learners usually do not repeat the lexical item, and Session 2, where the learner systematically repeats the lexical item before integrating it into his utterance.

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press

Table 3. Annotated multimodal transcription of a side sequence from Session 2

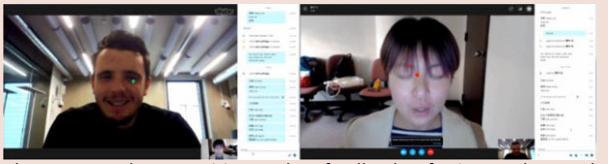
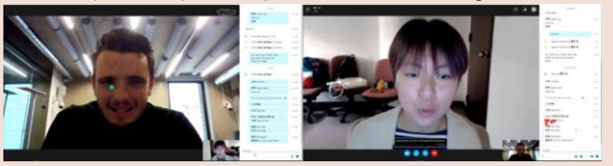
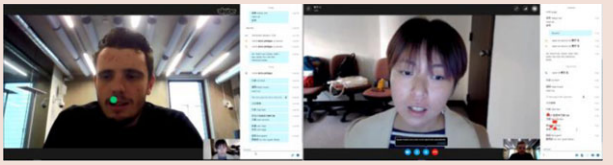
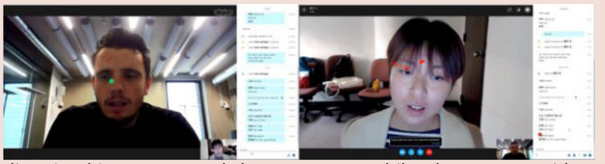
Turn	Participant and time code	Screenshots	Speech (translation)
1	Learner verbal 00:15:12.226 – 00:15:22.466	 The learner begins his utterance while looking at the chat window, which is open on both sides and has been used mainly by the tutor to annotate new words and sentences during the interaction. While the learner speaks, the tutor looks at his video feed and nods, providing positive feedback.  This head movement seems to draw the attention of the learner while he speaks, before redirecting his gaze at the chat window.  In the middle of his utterance, the learner faces a lexical gap 餐馆 (restaurant). This is visible through three joint behaviours: verbal hesitations with a slightly lengthened word 中国 (here "Chinese") and the filler 嗯 (em), a long intra-turn pause, and leaning sidewards to better look at the chat window. While the learner produces the filler 嗯, the tutor seems to be in tension: she raises her right hand and she has her mouth open, waiting for the learner to successfully finish his utterance.  The learner then continues his utterance producing an inappropriate formulation in Chinese while directing his gaze briefly toward the interlocutor, who changes her facial expression, lifting her eyebrows and almost biting her lips. This is negative feedback, but it is not perceived by the learner, who is now looking at the chat window. The learner smiles while leaning sidewards and looking for the right word in the chat window (餐馆).  At this moment, we can see that the tutor sees the learner looking for the word in the chat window, and she also looks at the word on her screen, which may be interpreted as an example of joint attention.	在法国嗯有很多+中国:嗯(0.9) 嗯有很多吃的嗯+餐馆 (in France em there are lots of Chinese: em (0.9) em there are lots of to be eaten/to eat em+restaurants)

(*Continued*)

[5]On the left is the tutor's screen and fixations (green dots); on the right, the learner's ones (red dots).

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press

Table 3. (*Continued*)

Turn	Participant and time code	Screenshots	Speech (translation)
2	Tutor verbal 00:15:22.579 – 00:15:26.980	 The tutor produces positive spoken feedback referring to the word.  Then she repairs the learner's utterance, with a short intra-turn pause to detach the nominal segment 中国饭馆(Chinese restaurants). Turn 2 is also an opportunity for the tutor to repair the tones and the learner's pronunciation more generally. While the tutor speaks, the learner initially looks at her and smiles, then looks back again at the word in the chat window.	对 + 在法国有很多+中国饭馆 (correct + in France there are lots of + Chinese restaurants)
3	Learner verbal 00:15:27.390 – 00:15:32.743	 When the tutor ends her turn, the learner nods and in turn 3 he once again repeats the utterance smoothly, without the hesitations present in his previous turn, while looking at the chat window, then  directing his gaze toward the tutor. Meanwhile, the tutor provides positive feedback by nodding while she mirrors the learner's uttering the sentence with lip movements, which is unnoticed by the learner who focuses on the tutor's eyes.	在在法国有很多中国饭馆 (in in France there are lots of Chinese restaurants)

4.3 Sequences of normative evaluation

Our second example comes from Session 2 (Table 3), during which we recorded gaze data for both tutor and learner. The side sequence in question emerges in the middle of a larger sequence about Chinese food and the presence of Chinese restaurants in France. It is an instance of IRF exchange, in which the question asked by the tutor in the initiation phase is less aimed at obtaining new factual information than at gaining an understanding of whether the learner is capable of producing the expected answer. We do not reproduce the first part of the exchange, in which the tutor asked if there are lots of Chinese restaurants in France and the learner answered yes. Rather, the excerpt starts after the tutor asked the learner to repeat himself.[6]

This example of SEN is highly representative of the instances we found in our corpus; everything we note occurred in several other sessions, apart from the episode of joint attention.

[6]The video is available at https://amupod.univ-amu.fr/video/20638-exemple-2mp4/

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press

Moreover, some of the multimodal strategies at play in it are the same as those being used in SPAs. The most obvious of these common patterns is the use of video as an affordance for the tutor to understand what a learner is doing, and possibly to leave the floor to the interlocutor(s) for maximum autonomy in expression. This is particularly visible at point 5 of turn 1 (Table 3), where the learner's ostension of leaning sideways leads to the tutor's inference that he is looking into the chat window for the word he needs for his utterance, which results in the episode of joint attention. In other words, her awareness of the learner's screen and the perception of the learner's ostensions lead the tutor to an inference concerning his experience, which we can interpret from the recording of the tutor's eye movements. Moreover, since we have the learner's eye-tracking data as well, we can confirm the tutor's inference.

On the learner side, video is used also to wait for feedback after turn completion. As shown in the second example, feedback can be positive or negative – in the latter case, usually without overt verbal indication. Indeed, in our data, verbal feedback is mostly positive, possibly followed by repair, as in the second example. The other common pattern is the use of the audio channel as an affordance to deploy the multimodality of speech, for instance, with the use of short intra-turn pauses by the tutor to draw learners' attention to specific lexical items.

In our data, SENs are usually shorter than SPAs. The tutors do not interrupt learners' turns and leave them the floor, even when there are long intra-turn pauses. The tutors signal they understand through ostensions using configurations of behaviours through audio and video, such as paraverbal sounds and head movements, or gesture and facial expressions as above.

5. Discussion and conclusion

Our first and main research question concerned the potentialities and limits of eye-tracking technology for the study of interlocutors' perception and use of multimodality in webconferencing. On the whole, our analysis confirms the utility of adopting a holistic approach to the study of multimodality during webconferencing-based language learning – more particularly, an ecological approach informed by multimodal CA for micro-analysis of communicative behaviours on the one hand, and by relevance theory to interpret the cognitive dimension on the other hand. Indeed, the methodological combination of CA tools with eye-tracking data enabled us to gain insight into not only the interlocutors' orientations (Mondada, 2016) but also the cognitive dimensions of intentions and inferences rendered through sequential multimodal ostensions. In other words, eye-tracking data enriched our ecological analytical approach focusing on the co-construction of meaning and social actions in webconferencing through cycles of action-perception-interpretation (van Lier, 2004). More precisely, the eye-tracking technique enriched such observations by providing a window on the cognitive management of graphic and visual affordances during interaction. Two different dynamics emerging from our analysis show this contribution. The first is the link between interlocutors' orientations and the results of gaze analysis. Eye tracking provides evidence that posture and proxemics in relation to the screen are perceived and used as interactional gestures, and allow interlocutors to appreciate one another's moment-to-moment engagement in interaction. This finding calls into question pedagogical recommendations that tutors look directly into the webcam to give learners the impression of being looked at in the eye (Develotte *et al.*, 2010). Instead, we show that when an interlocutor detaches his gaze from the screen, he ceases to be able to efficiently manage the interaction based on the ostensions through the video affordance. The second dynamic illustrative of the contribution of eye-tracking data is that such data can be instrumental in identifying and analysing instances of joint attention (when data is obtained from both sides of an interaction). Our second example indicates that the tutor was able to imagine what was on the learner's screen and adopt the learner's perspective. We suggest that this will be especially useful in phases of a tutoring session where the tutor mediation is key to directing learners' attention to particular

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press

parts of the screen, such as giving instructions, accompanying reading comprehension, analysing visual elements, and resolving technical issues.

As for its limits, eye tracking provides only partial information about interactions in webconferencing, which is not fully interpretable without other sources of data, especially audio and video recordings and dynamic screen captures. Therefore, it is less likely to be a useful stand-alone tool for analysis. A further limitation was imposed by our specific choice of eye-tracking tool, which sometimes provides approximations of fixations that were only accurate to 0.5 cm. More precise data could be gathered, but only using tools that were much more expensive and/or that restricted the interlocutors' head movements, thus reducing ecological validity. The third and final limitation concerns the difficulty of interpreting the meanings of gaze behaviours. In fact, even if eye-tracking data can provide fully accurate information about fixations on elements of the screen, this does not constitute a direct window on interlocutors' intentions. This limit may be partly overcome via stimulated recall, possibly using screen recordings with eye-tracking data superimposed on them. This method, like any other kind of retrospective explanation, is of course subject to cognitive bias (Mercier & Sperber, 2017).

Our study also yielded an answer to our secondary research question: some multimodal conversational patterns were shared between our two groups, while others were specific to one or the other. In particular, our analysis shows that video became an affordance for, on the one hand, the tutor to interpret learners' orientations and gazes as signs of engagement, and, on the other hand, for the learner(s) to solicit and interpret tutors' (conversational) feedback. However, given that this was a case study based on a very limited dataset, our findings regarding this question are at best only preliminary. Future work on much larger datasets with similar types of participants will help us understand both the possible variations and the common features of multimodal interaction. Ideally, any such extension should also include longitudinal data, which would shed new light on how communicative behaviour in general, and gaze in particular, evolves over time.

Supplementary material. To view supplementary material referred to in this article, please visit https://doi.org/10.1017/S0958344022000076

Acknowledgments. This study was possible thanks to the support of different research engineers and technicians at different stages of data collection and treatment. We would like to express our gratitude to Fabrice Cauchard of the H2C2 platform, Alain Ghio and Antonio Serrato from platform Centre d'Etudes sur la Parole at the Laboratoire Parole & Langage, and Christelle Zielinski from the Centre de Ressources Expérimentation of the Institute of Language, Communication and the Brain. We would also like to thank Xia Wang, Li Tang, Anqi Xu, Eugene YC Wong and KT Tong at the CBS Speech and Language Sciences (SLS) Lab in the Hong Kong Polytechnic University for their technical support. Our thanks also goes to the three anonymous reviewers of the first versions of this paper for their interesting comments and for letting us improve this article. This paper is part of a research project funded by the French ANR (https://anr.fr/Projet-ANR-18-CE28-0011).

Ethical statement and competing interests. This study was approved by the research review board of the Hong Kong Polytechnic University prior to the beginning of data collection (HSEARS20170926001) and conducted in accordance with the board's ethical guidelines. The guidelines of Laboratoire Parole & Langage were followed in the collection of data in France, and the French participants signed a written consent to be recorded. All participants received an explanation of the study and its procedures and gave their informed consent prior to its commencement. The authors declare no competing interests.

References

Bavelas, J. B., Chovil, N., Coates, L. & Roe, L. (1995) Gestures specialized for dialogue. *Personality and Social Psychology Bulletin*, 21(4): 394–405. https://doi.org/10.1177/0146167295214010

Bezemer, J. & Kress, G. (2016) *Multimodality, learning and communication: A social semiotic frame*. London: Routledge. https://doi.org/10.4324/9781315687537

Blin, F. (2016) The theory of affordances. In Caws, C. & Hamel, M.-J. (eds.), *Language-learner computer interactions. Theory, methodology and CALL applications*. Amsterdam: John Benjamins, 41–64. https://doi.org/10.1075/lsse.2.03bli

Bronfenbrenner, U. (1979) *The ecology of human development: Experiments by nature and design*. Cambridge MA: Harvard University Press.

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press

Cappellini, M. (2016) Roles and scaffolding in teletandem interactions: A study of the relations between the sociocultural and the language learning dimensions in a French–Chinese teletandem. *Innovation in Language Learning and Teaching*, 10(1): 6–20. https://doi.org/10.1080/17501229.2016.1134859

Cappellini, M. (2021) Une approche multimodale intégrant l'oculométrie pour l'étude des interactions télécollaboratives par visioconférence. *Les Cahiers de l'ASDIFLE*, 31: 99–120.

Cappellini, M. & Azaoui, B. (2017) Sequences of normative evaluation in two telecollaboration projects: A comparative study of multimodal feedback through desktop videoconference. *Language Learning in Higher Education*, 7(1): 55–80. https://doi.org/10.1515/cercles-2017-0002

Cappellini, M. & Hsu, Y.-Y. (2018) Ce que l'oculométrie peut apporter dans une approche écologique des échanges en ligne. Une discussion épistémologique et une étude de cas. In Dejean-Thircuir, C., Mangenot, F., Nissen, E. & Soubrié, T. (eds.), *Actes du colloque Échanger Pour Apprendre en Ligne 2018*. http://hal.univ-grenoble-alpes.fr/EPAL/hal-02023002

Cappellini, M. & Hsu, Y.-Y. (2020) When future teachers meet real learners through telecollaboration: An experiential approach to learn how to teach languages online. *Journal of Virtual Exchange*, 3: 1–11. https://doi.org/10.21827/jve.3.35751

Cohen, C. (2017) Former à l'enseignement en ligne. In Guichon, N. & Tellier, M. (eds.), *Enseigner l'oral en ligne: Une approche multimodale*. Paris: Didier, 215–242.

Conklin, K., Pellicer-Sánchez, A. & Carrol, G. (2018) *Eye-tracking: A guide for applied linguistics research*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108233279

Council of Europe (2001) *Common European framework of reference for languages: Learning, teaching, assessment*. Strasbourg: Language Policy Unit. http://www.coe.int/t/dg4/linguistic/source/framework_en.pdf

de Pietro, J.-F., Matthey, M. & Py, B. (1989) Acquisition et contrat didactique: les séquences potentiellement acquisitionnelles dans la conversation exolingue. In Weil, D. & Fugier, H. (eds.), *Actes du troisième Colloque Régional de Linguistique*. Strasbourg: Université des Sciences Humaines et Université Louis Pasteur, 99–124.

Develotte, C., Guichon, N. & Vincent, C. (2010) The use of the webcam for teaching a foreign language in a desktop videoconferencing environment. *ReCALL*, 22(3): 293–312. https://doi.org/10.1017/S0958344010000170

Develotte, C., Kern, R. & Lamy, M.-N. (eds.) (2011) *Décrire la conversation en ligne: Le face à face distanciel*. Lyon: ENS Editions.

Dooly, M. (2010) The teacher 2.0. In Guth, S. & Helm, F. (eds.), *Telecollaboration 2.0: Language, literacies and intercultural learning in the 21st century*. Bern: Peter Lang, 277–303.

Dooly, M. (2016) 'Please remove your avatar from my personal space': Competences of the telecollaboratively efficient person. In O'Dowd, R. & Lewis, T. (eds.), *Online intercultural exchange: Policy, pedagogy, practice*. Abingdon: Routledge, 192–208.

Fuchs, C., Hauck, M. & Müller-Hartmann, A. (2012) Promoting learner autonomy through multiliteracy skills development in cross-institutional exchanges. *Language Learning & Technology*, 16(3): 82–102.

Gibson, J. J. (1979) *The ecological approach to visual perception*. London: Lawrence Erlbaum Associates.

Gillespie, J. (2020) CALL research: Where are we now? *ReCALL*, 32(2): 127–144. https://doi.org/10.1017/S0958344020000051

Goodwin, C. (2000) Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32(10): 1489–1522. https://doi.org/10.1016/S0378-2166(99)00096-X

Guichon, N. (2012) *Vers l'intégration des TIC dans l'enseignement des langues*. Paris: Didier.

Guichon, N. & Cohen, C. (2016) Multimodality and CALL. In Farr, F. & Murray L. (eds.), *The Routledge handbook of language learning and technology*. London: Routledge, 509–521.

Guichon, N. & Tellier, M. (eds.) (2017) *Enseigner l'oral en ligne: Une approche multimodale*. Paris: Didier.

Guichon, N. & Wigham, C. R. (2016) A semiotic perspective on webconferencing-supported language teaching. *ReCALL*, 28(1): 62–82. https://doi.org/10.1017/S0958344015000178

Gullberg, M. & Holmqvist, K. (2006) What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics & Cognition*, 14(1): 53–82. https://doi.org/10.1075/pc.14.1.05gul

Hampel, R. & Stickler, U. (2005) New skills for new classrooms: Training tutors to teach languages online. *Computer Assisted Language Learning*, 18(4): 311–326. https://doi.org/10.1080/09588220500335455

Hampel, R. & Stickler, U. (2012) The use of videoconference to support multimodal interaction in an online language classroom. *ReCALL*, 24(2): 116–137. https://doi.org/10.1017/S095834401200002X

Hampel, R. & Stickler, U. (eds.) (2015) *Developing online language teaching: Research-based pedagogies and reflective practices*. New York: Palgrave Macmillan.

Hauck, M. (2010) Telecollaboration: At the interface between multimodal and intercultural communicative competence. In Guth, S. & Helm, F. (eds.), *Telecollaboration 2.0: Language, literacies and intercultural learning in the 21st century*. Bern: Peter Lang, 219–244.

Helm, F. & Dooly, M. (2017) Challenges in transcribing multimodal data: A case study. *Language Learning & Technology*, 21(1): 166–185.

Holt, B. & Tellier, M. (2017) Conduire des explications lexicales. In Guichon, N. & Tellier, M. (eds.), *Enseigner l'oral en ligne*. Paris: Didier, 59–90.

Hubbard, P. (2017) An invitation to CALL: Foundations of computer-assisted language learning. In Son, J.-B. & Windeatt, S. (eds.), *Language teacher education and technology: Approaches and practices*. London: Bloomsbury, 153–168

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press

Jefferson, G. (1972) Side sequences. In Sudnow, D. (ed.), *Studies in social interaction*. New York: The Free Press, 294–338.
Kendon, A. (1982) The study of gesture: Some observations on its history. *Recherches sémiotiques/Semiotic Inquiry*, 2(1): 25–62.
Kessler, G. (2016) Technology standards for language teacher preparation. In Farr, F. & Murray, L. (eds.), *The Routledge handbook of language learning and technology*. Abingdon: Routledge, 57–70.
Kozar, O. (2016) Perceptions of webcam use by experienced online teachers and learners: A seeming disconnect between research and practice. *Computer Assisted Language Learning*, 29(4): 779–789. https://doi.org/10.1080/09588221.2015.1061021
Lamy, M.-N. & Hampel, R. (2007) *Online communication in language learning and teaching*. New York: Palgrave Macmillan. https://doi.org/10.1057/9780230592681
Mercier, H. & Sperber, D. (2017) *The enigma of reason*. London: Penguin.
Mondada, L. (2016) Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics*, 20(3): 336–366. https://doi.org/10.1111/josl.1_12177
Mondada, L. & Pekarek Doehler, S. (2004) Second language acquisition as situated practice: Task accomplishment in the French second language classroom. *The Modern Language Journal*, 88(4): 501–518. https://doi.org/10.1111/j.0026-7902.2004.t01-15-.x
O'Dowd, R. (2018) From telecollaboration to virtual exchange: State-of-the-art and the role of UNICollaboration in moving forward. *Journal of Virtual Exchange*, 1: 1–23. https://doi.org/10.14705/rpnet.2018.jve.1
O'Rourke, B. (2012) Using eye-tracking to investigate gaze behavior in synchronous computer-mediated communication for language learning. In Dooly, M. & O'Dowd, R. (eds.), *Researching online foreign language interaction and exchange: Theories, methods and challenges*. Bern: Peter Lang, 305–342.
Pekarek Doehler, S. (ed.) (2000) Approches interactionnistes de l'acquisition des langues étrangères: concepts, recherches, perspectives. *Acquisition et Interaction en Langue Étrangère*, 12: 1–19. https://doi.org/10.4000/aile.934
Py, B. (2000) La construction interactive de la norme comme pratique et comme représentation. *Acquisition et Interaction en Langue Étrangère*, 12: 77–97. https://doi.org/10.4000/aile.1464
Rivens Mompean, A. & Cappellini, M. (2015) Teletandem as a complex learning environment: Looking for a model. *DELTA*, 31(3): 633–663. https://doi.org/10.1590/0102-4450430446379623426
Satar, H. M. (2016) Meaning-making in online language learner interactions via desktop videoconferencing. *ReCALL*, 28(3): 305–325. https://doi.org/10.1017/S0958344016000100
Satar, H. M. & Wigham, C. R. (2017) Multimodal instruction-giving practices in webconferencing-supported language teaching. *System*, 70: 63–80. https://doi.org/10.1016/j.system.2017.09.002
Satar, H. M. & Wigham, C. R. (2020) Delivering task instructions in multimodal synchronous online language teaching. *Alsic*, 23(1). https://doi.org/10.4000/alsic.4571
Sert, O. & Balaman, U. (2018) Orientations to negotiated language and task rules in online L2 interaction. *ReCALL*, 30(3): 355–374. https://doi.org/10.1017/S0958344017000325
Sinclair, J. M. & Coulthard, R. M. (1975) *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford: Oxford University Press.
Sloetjes, H. & Wittenburg, P. (2008) Annotation by category - ELAN and ISO DCR. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S. & Tapias, D. (eds.), *Proceedings of the Sixth International Conference on Language Resources and Evaluation*. Marrakech: European Language Resources Association, 816–820.
Sperber, D. & Wilson, D. (1986) *Relevance: Communication and cognition*. Oxford: Blackwell Publishing.
Stickler, U., Smith, B. & Shi, L. (2016) Using eye-tracking technology to explore online learner interactions. In Caws, C. & Hamel, M.-J. (eds.), *Language-learner computer interactions: Theory, methodology and CALL applications*. Amsterdam: John Benjamins, 163–186. https://doi.org/10.1075/lsse.2.08sti
van Lier, L. (2004) *The ecology and semiotics of language learning: A sociocultural perspective*. Dordrecht: Kluwer Academic. https://doi.org/10.1007/1-4020-7912-5
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. https://doi.org/10.1515/cercles-2017-0001
Yamada, M. & Akahori, K. (2010) Awareness and performance through self- and partner's image in videoconferencing. *CALICO Journal*, 27(1): 1–25. https://doi.org/10.11139/cj.27.1.1-25

Author ORCiD. Marco Cappellini, https://orcid.org/0000-0002-2086-061X
Author ORCiD. Yu-Yin Hsu, https://orcid.org/0000-0003-4087-4995

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press

About the authors

Marco Cappellini is an associate professor in didactics of foreign languages at Aix-Marseille University. His research interests include tandem learning, interactionist approaches to foreign language (FL) learning, FL teaching and learning through CMC, teacher education and the integration of ICT in FL education, and learner autonomy.

Yu-Yin Hsu is an assistant professor in the Department of Chinese and Bilingual Studies at the Hong Kong Polytechnic University. Her research interests include technology application in FL teaching and learning, FL teacher training, psycholinguistic language processing, and theoretical linguistics.

https://doi.org/10.1017/S0958344022000076 Published online by Cambridge University Press