ReCALL (2021), 33: 3, 195–213
doi:10.1017/S0958344021000070

RESEARCH ARTICLE

Multimodal (inter)action analysis of task instructions in language teaching via videoconferencing: A case study

Ciara R. Wigham
Université Clermont Auvergne, France (ciara.wigham@uca.fr)

Müge Satar
Newcastle University, United Kingdom (muge.satar@newcastle.ac.uk)

Abstract

Online language teaching is gaining momentum worldwide and an expanding body of research analyses online pedagogical interactions. However, few studies have explored experienced online teachers' practices in videoconferencing particularly while giving instructions, which are key to success in task-based language teaching (Markee, 2015). Adopting multimodal (inter)action analysis (Norris, 2004, 2019) to investigate the multimodal construction of instructions in a single case study, we examine instruction-giving as a social practice demonstrated in a specific site of engagement (a synchronous online lesson recorded for research purposes). Drawing on the higher-level actions (instruction-giving fragments) we have identified elsewhere (Satar & Wigham, 2020), in this paper we analyse the lower-level actions (modes) that comprise these higher-level actions, specifically focusing on the print mode (task resource sheets, URLs, text chat, and online collaborative writing spaces) wherein certain higher-level actions become frozen. Our findings are unique in depicting the modal complexity of sharing task resources in synchronous online teaching due to semiotic misalignment and semiotic lag that precludes the establishment of a completely shared interactional space. We observe gaze shifts as the sole indicator for learners that the teacher is multitasking between different higher-level actions. Further research is needed to fully understand the interactional features of online language teaching via videoconferencing to inform teacher training policy and practice.

Keywords: multimodal (inter)action analysis; task-based language teaching; videoconferencing; instruction-giving practices; print mode; frozen actions

1. Introduction

Online language learning via videoconferencing is an increasingly important area within the computer-assisted language learning (CALL) field. As online language schools grow in number, offering flexible learning and teaching delivery methods, and education institutions including schools and universities turn to videoconferencing in hybrid and distance language learning programmes, the field needs to better understand successful online language teachers' semiotic pedagogical skills (Develotte, Guichon & Vincent, 2010).

An expanding body of research analyses synchronous multimodal pedagogical interactions via videoconferencing and addresses semiotic modes used in online environments for specific purposes (Guichon & Drissi, 2008; Hampel & Stickler, 2012; Lee, Hampel & Kukulska-Hulme, 2019), discourse types (Guichon, Bétrancourt & Prié, 2012; Wang, 2006), the effects of the webcam on multimodal interactive learning (Codreanu & Celik, 2012; Cohen & Wigham, 2019;

Cite this article: Wigham, C.R. & Satar, M. (2021). Multimodal (inter)action analysis of task instructions in language teaching via videoconferencing: A case study. *ReCALL* 33(3): 195–213. https://doi.org/10.1017/S0958344021000070

© The Author(s), 2021. Published by Cambridge University Press on behalf of European Association for Computer Assisted Language Learning.

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press

Develotte *et al*., 2010; Guichon & Wigham, 2016; Kern, 2014), and social presence (Guichon & Cohen, 2014; Satar, 2013, 2016). Although most studies investigated pedagogical interactions within a task-based approach, research has not specifically focused on instruction delivery despite this potentially being a key pedagogical competence for task success (Watson Todd, Chaiyasuk & Tantisawetrat, 2008). Furthermore, published research has focused upon trainee teachers (Cappellini & Combe, 2017; Satar & Wigham, 2017), with research sites being teacher-training courses led by the researchers. Few studies concern experienced online teachers.

This case study examines how experienced online language teachers harness the print mode (resource sheets, URLs, text chat, Google Docs [online document hereafter]) in instruction-giving practices. Adopting multimodal (inter)action analysis (Norris, 2004, 2019), we explore the video-conferencing interface as a site of engagement or "window opened up through practices that make concrete mediated actions possible" (Scollon, 1998, in Norris, 2019: 120). We then examine the print mode and its role in mediated actions, and then analyse modal configurations and density in the teacher's interactional space, indicating level of attention/awareness for different higher-level actions as the print mode is employed. We address three research questions:

1. How does the site of engagement demonstrate any intersection of different social practices?
2. What is the role of the print mode in achieving the higher-level action of instruction-giving?
3. How does the teacher's modal configuration and density vary when the print mode is utilised?

2. Theoretical background

2.1 Online teaching competencies

Synchronous online language learning is a semio-pedagogical activity that requires the teacher's use of appropriate socio-affective, pedagogical, semiotic, and technological competencies to create and maintain optimal learning conditions to favour language skill development. Guichon (2009) outlines these skills, describing socio-affective regulation as the ability to establish and maintain a relationship with learners. Pedagogical regulation requires expert knowledge of the target language, the ability to design learning scenarios adapted to online learning, and propose "clear and concise instructions, providing positive and negative feedback and deploying an array of strategies to facilitate second-language learning" (p. 169). Finally, semiotic and technological competencies relate to the online teacher adopting the different communication tools available in the technological environment that are appropriate to the learning scenario and his or her ability to manage the ensuing interactions using adequate semiotic resources in order to optimize learning potential and maintain learners' attention (Ricci Bitti & Garotti, 2011). This latter competency is crucial, as teaching via videoconferencing allows teachers to communicate through an array of modes, including spoken language, print, posture, gesture, head movement, gaze, and proxemics, and thus demands that teachers develop *critical semiotic awareness* (Guichon, 2013; Kern, 2015) considered as "a high level of consciousness regarding all the information they are conveying when they interact with their learners" (Guichon & Wigham, 2016: 67).

2.2 Online teaching and the use of print mode

One affordance of online platforms utilised for teaching synchronously is the availability of text chat. Text chat affords the potential to enhance teaching using the spoken mode with written language and visuals (Meskill & Anthony, 2010). This potential enables the language teacher to capitalise on the multimodality of the teaching medium, for example, by offering feedback without interrupting the learners, encouraging learners who are unwilling to participate orally to contribute in writing, checking and confirming vocabulary upon communication breakdown, and to compensate for the absence of a blackboard (Hampel & Stickler, 2005, 2012; Wigham & Chanier, 2015).

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press

Investigating synchronous language teachers' areas of attention on the screen in an audio-conferencing platform using heat maps and gaze plots recorded in an eye-tracking study, Shi, Stickler and Lloyd (2017) categorised text chat as part of the technical areas of interest "where participants activate communication modes" (p. 217), and demonstrated teachers' monitoring of the text chat box during all lesson stages (i.e. opening, teaching, and closing). Concerning video-conferencing (i.e. Skype-based English lessons), Kozar (2016) studied how experienced online teachers employed text chat to especially serve pedagogical and organisational functions. Outlining both benefits and challenges, Kozar concluded that although text chat serves predominantly pedagogical and organisation functions, learners incorporated "a text message if: (a) it introduced new vocabulary, and (b) if it was produced bimodally (speaking and typing)" (p. 231).

The ability to employ the print mode is, thus, crucial for online teachers who should be "comfortable in using computer-mediated communication (CMC) technologies which include text, audio, and video conferencing" (Compton, 2009: 83). This brief overview suggesting online teachers should skilfully utilise the print mode, particularly text chat, in synchronous teaching attests to the need to further explore modal configuration and use in interaction.

2.3 Delivering task instructions

The context for this study is task-based language teaching (TBLT) in multimodal synchronous online language teaching. TBLT encourages meaning-focused language use to achieve an outcome, and a task involves, first, some input for learners to process and use in their interactions, and, second, instructions explaining the outcome to be achieved through interaction, as well as how to achieve it (Ellis, 2000). Drawing on Breen (1987), although teachers can plan task instructions as part of their designed task (instructions-as-workplan), the actual instructions they deliver in interaction with the learners might differ (instructions-as-process). Investigating this perspective in which teachers give instructions verbally during face-to-face lessons, Markee (2015) identified six key instruction-giving fragments, including explanation of what resources learners will need during task accomplishment. In Satar and Wigham (2020), we analysed experienced language teachers' pedagogical instructions to build on Markee's (2015) fragments. However, given our analytical lens, multimodal (inter)action analysis (Norris, 2004), which scrutinises mediated higher- and lower-level actions, in our work Markee's (2015) fragments constitute various higher-level actions the teachers engage in as part of their instruction-giving practices, while the modes they employ in doing so form the lower-level actions. Satar and Wigham (2020) described that explanation of resources can present challenges in the online teaching context due to a lack of shared physical interactional space and artefacts (e.g. task handouts). Rather, teachers need to navigate, for example, sharing URL links to resource websites or shared online collaborative word processors via text chat, and the distribution of text documents to learners via email or text chat. Satar and Wigham (2020) reported that the higher-level action labelled *managing resources* took about a quarter of the instruction-giving sequence in two of the three teachers' lessons examined, and observed the following sub-categories: *sending the resource*, *allocating the resource*, *receiving the resource*, *opening the resource*, *confirming access to the correct resource*, *describing the content of the resource*, and *reading the resource*. This paper focuses on this higher-level action of *managing resources*.

Within the online language teaching literature, two studies to date have addressed instruction-giving practices, both focusing on trainee teachers. Comparing written instructions in an asynchronous teaching platform with verbal instructions in multimodal synchronous teaching, Cappellini and Combe (2017) reported that trainee teachers experienced unexpected challenges concerning managing the digital task resources, such as having to instruct the learners on how to download a document, especially in the synchronous platform. Satar and Wigham (2017) addressed the multimodal resources trainee teachers employed while delivering task instructions,

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press

and demonstrated the use of text chat in introducing key vocabulary items in both spoken and written language (dual modality).

Given this background in online language teaching, we now explain the main concepts from multimodal (inter)action analysis that underpin our study.

2.4 Multimodal (inter)action analysis

Multimodal (inter)action analysis (Norris, 2004, 2019, 2020) is "a holistic analytical framework that understands the multiple modes in (inter)action as all together building one system of communication" (Norris & Pirini, 2016: 24). Within this framework, all actions are considered interactions between social actors and other social actors, objects, or the environment. It offers a broad sociocultural perspective that is compatible with detailed microanalysis of interaction, allowing the researcher to address the macro-, intermediate, and microanalyses as required (Norris & Pirini, 2016). We now explain the key analytical tools of multimodal (inter)action analysis.

2.4.1 Mediated actions

In multimodal (inter)action analysis, the unit of analysis is the mediated action "defined as a social actor acting with/through mediational means" (Norris & Pirini, 2016: 24). Mediated actions are investigated as higher-level and lower-level actions: while lower-level actions are "the smallest interactional meaning unit" (Norris, 2004: 11), higher-level actions are "a chain of lower-level actions, with an opening and a closing" (Norris & Pirini, 2016: 25). For example, a conversation is a higher-level action that constitutes lower-level actions of, for instance, gaze shifts (gaze mode) or utterances (spoken language mode). Higher-level actions can also be embedded in various other levels of higher actions: a conversation among friends can occur within the higher-level action of walking.

Frozen actions are the third type of actions. They are higher-level actions performed by social actors at an earlier time and that are now entailed in disembodied modes (e.g. printed material, layout of an environment) or material objects (Norris, 2004). For instance, Norris and Makboon (2015) illustrate the higher-level action of raising children as embedded in a self-made photo calendar present in a web-designer's office and telling of her identity as a working mother.

2.4.2 Site of engagement

A site of engagement is the "'real time window' opened through the intersection of social practices and mediational means that enables a mediated action to occur" (Norris & Jones, 2005: 139) wherein higher-level, lower-level, and frozen actions happen. It comprises the social actors and the mediated actions in which the actors are engaged and the frozen actions entailed in material objects within the interactional setting. In our data, the observed frozen actions are entailed in electronic resources in the print mode.

2.4.3 Communication modes and the print mode

Norris (2004) identifies nine communication modes: spoken language, proxemics, posture, gesture, head movement, gaze, layout, music, and print. Layout is a disembodied mode, whereas music and print can be both embodied and disembodied. Norris suggests that print is a visual mode and "refers to written texts, including the language, the medium, the typography, and the content ... [and] images in the printed media" (2004: 44). When people comment on the content of printed media (such as magazines, shopping lists), print is a disembodied mode: the act of writing is frozen as a higher-level action. However, when the mode is employed to

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press

express thoughts/feelings (e.g. by writing or manipulating shopping lists or creating images), print becomes an embodied mode.[1]

2.4.4 Modal configuration, modal density, and attention/awareness continuum

Before presenting our analyses, notions of modal configuration and modal density require explanation. Modal configuration (hierarchical organisation of lower-level actions) allows the analysis of a higher-level action in terms of the (chains of) lower-level actions (e.g. gesture units, utterances, printed materials) that constitute it and their relationships (Norris, 2004). In this process "[t]he lower-level actions that are most important to the meaning produced are defined as most important to the construction of the higher-level action" (Norris & Pirini, 2016: 26).

Following the analysis of a higher-level action's modal configuration, its modal density can be mapped out. Modal density can be measured through *intensity* and *complexity*. When a mode is absolutely necessary for the completion of a higher-level action, it carries *high modal intensity*. When the mode is discontinued, if it changes the higher-level action only slightly, then it has *medium modal intensity*; when it does not change the higher-level action, it carries *low modal intensity* (Norris, 2004). Communication modes take on particular intensity in specific interactions. For example, when speaking on the phone the spoken mode is strongest. Even though the speaker might gesture or doodle as she or he speaks, the gesture and print modes carry less intensity and, if discontinued, do not change the higher-level action of partaking in a phone conversation. However, if another person enters the speaker's physical space and points at a clock, and in turn the speaker ends the phone call, the gesture mode carries high intensity: when discontinued, it leads to a change in higher-level action.

Modal complexity, diversely, "refers to the interplay of numerous communicative modes that make the construction of a higher-level action possible … [when the modes are] intricately intertwined" (Norris, 2004: 87). For example, for the higher-level action of building a piece of flat-pack furniture, the print mode in the instruction booklet, the layout mode with regard to the positioning of different components on the floor, and a person's posture towards certain components all enable the higher-level action.

Therefore, any given higher-level action can be modally dense either when (a) a mode plays a central role or (b) multiple modes are inextricably linked to produce the action.

Modal density is also a measure of social actors' phenomenological attention/awareness (Norris & Pirini, 2016). Social actors can pay "various levels of attention … to simultaneously performed higher-level mediated actions" (Norris, 2019: 247). The modal density of these actions determines their relative position in social actors' attention/awareness continuum: foreground, mid-ground, and background. When the higher-level action has high modal intensity, the social actor is more focused upon that action (Norris & Pirini, 2016), which is foregrounded in his/her attention/awareness.

3. Methods

This paper stems from a larger study investigating experienced online language teachers' instruction-giving practices in English as a foreign language lessons via videoconferencing. In Satar and Wigham (2020), we identified higher-level actions related to instruction-giving that were employed in the lessons drawing on grounded theory principles (Strauss & Corbin, 1998) and demonstrated how instructions are achieved using multimodal (inter)action analysis (Norris, 2004). Here, we propose a revelatory single case study (Yin, 2018) of one teacher (Craig) who offered us multiple opportunities to observe and analyse the semio-pedagogical skills of an online language teacher while the print mode is used during an instruction-giving fragment: *managing resources*. The case study relies on the interaction data from one of the videoconferencing lessons. Although in our larger study we did collect interview data, we are unable to report on it here due to space restrictions.

[1]For a discussion about the boundaries between embodied and disembodied print mode, see Norris (2004: 45–47).

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press

3.1 Participants and context

Craig had been teaching English via audio- and videoconferencing since 2016. At the time of data collection (May 2018), he had completed 1,454 lessons on one specific platform and had received excellent feedback (5 stars out of 5) from almost 200 students. The data for this paper comes from a 60-minute Skype lesson with two female Turkish learners (Eda and Didem) studying English for academic purposes in a foundation year of a Turkish university. All participants gave informed consent.

The lesson was the second in a series of three and was designed around a convergent task (Mercer, Wegerif & Dawes, 1999) requiring collaborative work. The task (see Appendix in the supplementary material) comprised two microtasks: (1) an information exchange activity in which the learners were given two different gift packages and required to compare their information in order to make a decision about which to buy for a colleague's leaving present; and (2) a collaborative writing activity in which the learners jointly composed an email to their colleagues about their gift choice. The task was designed by the researchers and two resource sheets (see Appendix in the supplementary material) were provided to the teacher, yet he had flexibility in deciding how to use this task in the lesson and give instructions for it.

3.2 Operationalising multimodal (inter)action analysis for this study

The data set used comprises two 60-minute videos of the lesson. The first is a screen-capture recording taken by the teacher himself. The second is the screen-capture recording taken by one researcher who partook in the lesson as a non-participant observer. This gave access to one potential recipients' view of the lesson.

Initially, the verbatim data were transcribed using ELAN (Sloetjes & Wittenburg, 2008) before using the software to take notes about concepts of interest that emerged over several viewings of the lesson. Following Norris (2019), this open coding allowed us to identify instruction-giving sequences for which we demarcated all sequences considered as higher-level actions (see Satar & Wigham, 2020).

To address RQ1, concerning the site of engagement, a screenshot of the teacher's online environment was taken 29 minutes 38 seconds into the interaction. This moment was chosen as the lesson was underway and any initial technical issues had been resolved. It formed part of the instruction-giving sequence for the information exchange activity but was representative of the overall data: all higher-level actions in this social practice are evident.

Regarding our second research question, we chose two data pieces related to the role of print for microanalysis. Our site of engagement was thus narrowed to part of the instruction-giving sequence for the information exchange activity and an instruction sequence regarding the second microtask. The teacher's and researcher's screen-recording videos were aligned and imported side by side into the ELAN workspace. Both researchers watched this video data together in ELAN, discussing their observations and taking notes. Multimodal transcriptions of the microactions that participants performed in the modes of gesture, proxemics, gaze, posture, and head movement[2] allowed us to analyse the (inter)actions in detail and embed aspects of analysis. The analytical tools and units of analysis used to conduct the microanalyses related to RQ2 are mediated actions: lower, higher, and frozen actions (Norris, 2019).

Multimodal transcription allowed us to analyse modal configuration and density (RQ3). In identifying the hierarchical ordering of lower-level actions, our guiding principle was to discern "what is absolutely necessary to perform this very action and what is not", as well as modal density to determine the place of the higher-level actions within the social actors' attention/awareness (Norris, 2019: 246). The analytical tools of modal density, modal configurations, and the modal density foreground–background continuum of attention/awareness (Norris, 2019) were employed.

[2]See Satar and Wigham (2017) for an explanation of our multimodal transcription methods.

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press

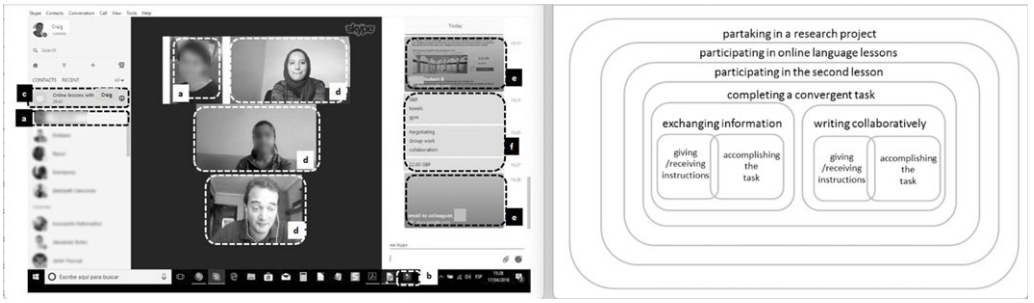


Figure 1. Left: site of engagement; Right: higher-level actions[3]

4. Analysis

4.1 RQ1: How does the site of engagement demonstrate any intersection of different social practices?

A screenshot (Figure 1, Left) allows us to analyse the different social practices constructed through the higher-level actions at play within the site of engagement (see section 2.4.2).

The site of engagement involves five social actors. Three actors (a volunteer teacher and two learners) responded to an advert to enlist participants for a research study (conducted by the other two actors: researchers) that comprises one large-scale action: partaking in a research project. This is evident in the print mode by the Skype avatar of one of the two principal investigators in the research project [a] and by a software icon in the toolbar for a screen-recording programme to capture the interaction data [b]. Although other social practices regarding data analysis are not visible (e.g. watching the videos, transcribing the mediated actions, writing up findings, etc.), Figure 1, Left, helps make sense of our part as researchers in relation to the data set as regards data collection.

The participants are engaged in a second large-scale action comprising participating in online language lessons via videoconferencing. This social practice is shown in the sidebar – the video-conferencing call is named "Online lessons with Craig" [c] – and through the presence of the teacher's and learners' webcam images [d].

The higher-level actions of completing language-learning tasks and also the higher-level actions of giving instructions for these tasks and of task accomplishment are evident in frozen actions entailed in the resources visible in the text chat window (i.e. in the disembodied print mode). They include task resource/instruction sheets [e] for both subtasks and lexical items offered by the teacher to learners as linguistic feedback during the accomplishment of the information exchange [f].

The site of engagement shows an intersection between the social practices of engaging in a research study and partaking in a teaching–learning activity constructed through the higher-level actions of participating in the second lesson, and particularly completing a convergent task (Figure 1, Right).

4.2 RQ2: What is the role of the print mode in achieving the higher-level action of instruction-giving?

We now examine the role of the print mode (text chat) in achieving the higher-level action of instruction-giving, and specifically managing resources, by examining Extract 1 (Figures 2 & 3) produced during the information-exchange activity (see section 3.1).

[3]Higher-resolution versions of the figures in this article can be viewed online as supplementary material.

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press

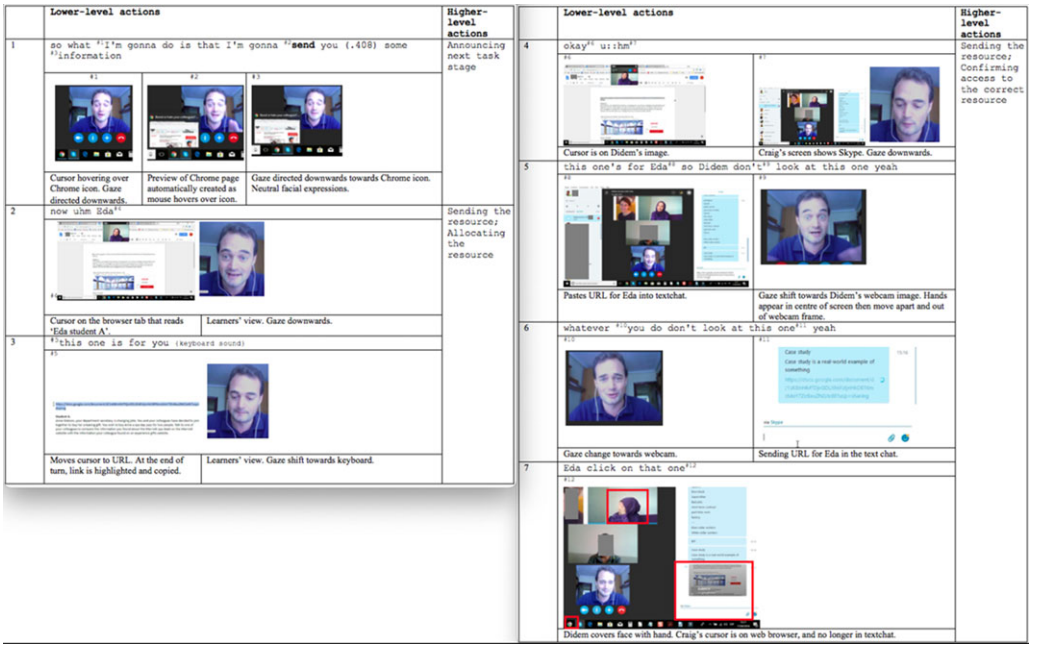


Figure 2. Left: Extract 1, lines 1–3; Right: Extract 1, lines 4–7

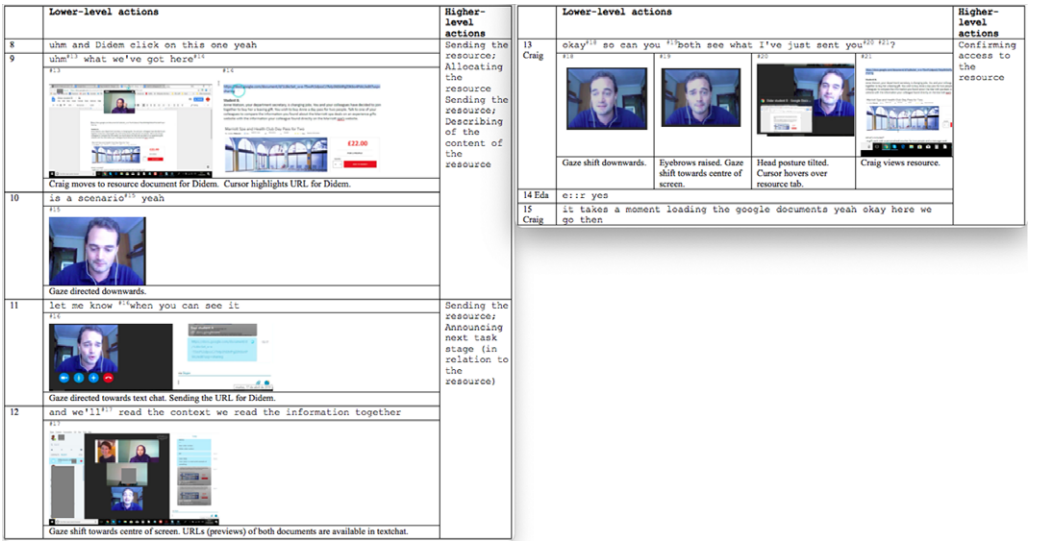


Figure 3. Left: Extract 1, lines 8–12; Right: Extract 1, lines 13–15

In line 1 (Figure 2), Craig begins *announcing the next task stage* with a gaze shift downwards (#1[4]). This is followed by Craig declaring what he will do next in the spoken language mode with emphasis on the word *send* (#2), with his gaze directed towards the learners, displaying an iconic gesture, and raised eyebrows acting as a beat gesture (for gesture types, see McNeill, 1992). As

[4]#n refers to the image number within the extract transcript being discussed.

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press

Craig tells the learners that he will send them some information, his gaze shifts towards the web browser, which displays an online newspaper in the current tab (#3).

In line 2, Craig begins to engage with two higher-level actions: sending the resource and allocating the resource to a specific learner. To allocate the resource, he first uses the learners' name, a vocative in the spoken language mode. Yet Craig's screen no longer portrays the webcam images of both learners. Rather, his focal point of attention, illustrated by his downward gaze (#4), is turned to a disembodied electronic resource in the print mode (an online document now displayed on the web browser, which includes the task information for student A, Eda).

In line 3, Craig highlights a URL in the resource sheet (#5), which he had generated and pasted into the document in preparation for the lesson. This document entails two frozen actions that are salient in this context, and which are the closest in time, place, and relevance to the current (inter) action: the researchers creating the resource sheet, and the teacher adding a URL to appropriate it for his own teaching practices. These frozen actions reinforce the site of engagement we described in section 4.1.

In line 4, Craig uses hesitation markers *okay u:hm* and downward gaze (#7) to signal an upcoming shift to the next instruction-giving fragment, *confirming access to the correct resource* while still *sending the resource*, as Craig places his mouse cursor on Didem's webcam image (#6) to activate Skype and access text chat (#7). In line 5, the print mode continues to be embodied while Craig pastes the URL previously copied (line 3) from the resource sheet into the text chat (#8). In lines 5–6, Craig reinforces allocation of the resource to one learner, Eda, by stating that the other learner, Didem, should not look at the URL he has sent. In doing so, his attention/awareness is on the Skype window, his gaze shifts upwards to focus on Didem's webcam image, his gestures become visible in his webcam frame, and he raises his eyebrows (#9). This information is repeated (line 6), and Craig's gaze moves to focus on his webcam as he uses the pronoun "you" to address Didem (#10) before the URL is sent on text chat (#11). In line 7, we see Didem's reaction: she covers her face with her hand and turns away from her screen (#12), demonstrating the success of this instruction-giving fragment.

In lines 8–12 (Figure 3, Left), Craig begins the higher-level action of *sending the resource* for the second learner, Didem. This requires a similar set of lower-level actions by embodying and disembodying the print mode (#13, #14, #16, and #17). Consequently, the simultaneous, multiple higher-level actions become frozen and available throughout the lesson.

An interesting point to note in lines 3–9 (Figures 2 and 3) is Craig's use of the demonstrative adjectives to refer to the URLs. In lines 3, 5, and 6 (#5, 6, 7, 8, 11), while selecting, copying, and pasting the first URL (embodied print mode), Craig refers to the link with the singular demonstrative adjective *this*. While embodying the print mode for the second URL, he again uses *this* (line 8) and adverb *here* (line 9); both signify closeness to the speaker in time and space. However, once the URL is sent, and thus disembodied (Figure 2, line 7, #12), Craig refers to it using the singular demonstrative adjective *that*, indicating distance from Craig in time and space.

Lines 13–15 (Figure 3, Right) demonstrate Craig's critical semiotic awareness as he engages with the higher-level action of confirming access to the resource in the spoken language mode by checking whether learners can see the resources, and announcing his awareness of a potential semiotic lag (line 15). Semiotic lag concerns a time delay or lag between the multimodal communication of one social actor (e.g. a gesture unit, a gaze shift, and electronic document) and its reception by another whereby it no longer forms an aggregate due to delays in online transmission or technical issues (e.g. microphone not working). For instance, one actor may send a text chat message in the print mode while simultaneously referring to it in the spoken language mode. However, due to semiotic lag, the receiving actor may receive the written message at a different moment in interaction than its temporal position in the sending actor's interactional space.

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press

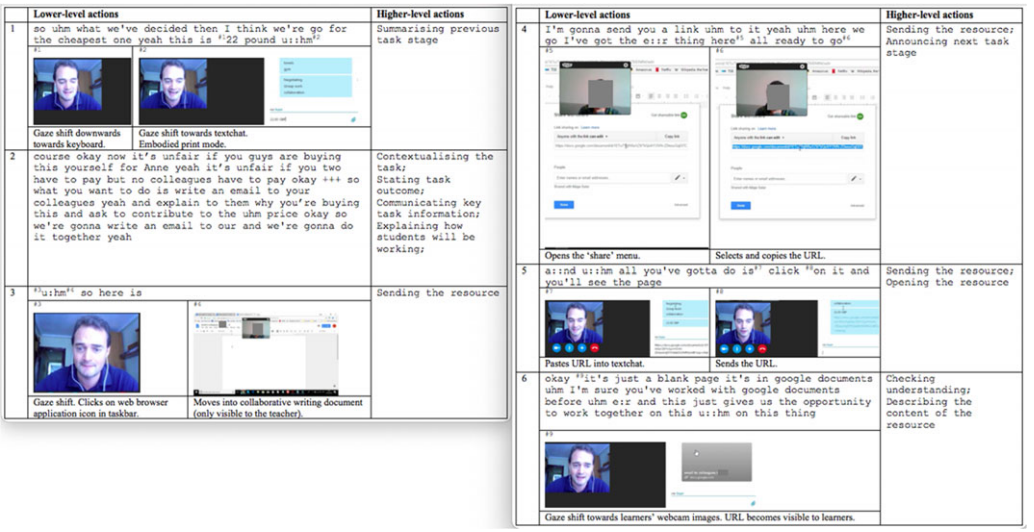


Figure 4. Left: Extract 2, lines 1–3; Right: Extract 2, lines 4–6

In line 14, Craig's confirmation check receives an affirmative response from one learner, before Craig announces verbally a shift into the next higher-level action (*here we go then*), which concludes the higher-level action of sending the resource.

In Extract 1, Craig achieves the higher-level actions of *sending* and *allocating* an electronic resource, and *confirming access to* and *describing the contents* of it by actively employing the print mode (text chat) alongside other available modes: gaze, gestures, facial expressions, and spoken language. Gaze shifts appear to be the most salient modal change signalling to learners when Craig is engaged with the technical aspects of sending the resource by embodying the print mode (downward gaze direction in #1–5, 7, 8, 15, 16, 18), and when he is not while directly addressing the learners (upward gaze direction in #9, 10, 17, 19, 20).

We now examine use of another communication technology (online document/print mode) to create content during instructions for the second microtask (see Appendix in the supplementary material), illustrating how the higher-level actions are created and become frozen within the document.

Extract 2 (Figures 4 & 5) begins with Craig summarising the decision made by the learners to choose the cheaper package in the first microtask. Craig types the price of this package in text chat and, similar to Extract 1, his embodiment of the print mode is signalled to the learners via downward gaze shifts (#1 and #2). *Summarising previous task stage* becomes a frozen action in the text chat as key information for the next stage. Craig then proceeds to verbally communicate the instructions (line 2), *contextualising the task* for the learners, *stating the task outcome* (an email), *giving key task information*, and *informing them how they will be working* (together), before initiating the higher-level action of *sending the resource* fragment (line 3). In line 3, Craig's gaze moves towards the web browser icon in the task bar at the bottom of his screen (#3), and once the pre-prepared online document is visible (#4), he proceeds to embody the resource in the print mode.

In line 4, Craig opens the menu in the resource to share the online document (#5), selects and copies the URL (#6), pastes it in text chat (#7), sends the URL (#8), and as it becomes visible to the learners (#9), the URL and the online document becomes disembodied again. During this time, his utterances show his critical semiotic awareness regarding *semiotic misalignment*. We define *semiotic misalignmen*t as a difference in the semiotic meaning-making resources available to the social actors involved in the interaction that distorts the shared interactional space. It can

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press

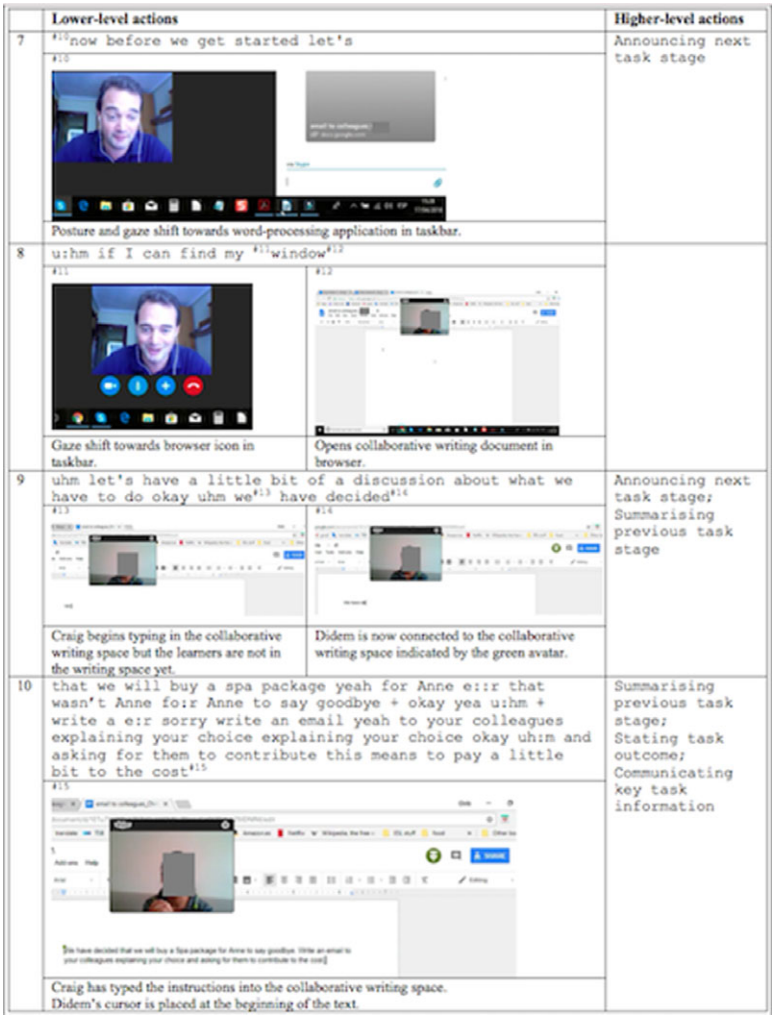


Figure 5. Extract 2, lines 7–10

be due to two factors. First, a difference in how each social actor views the interactional space where both/all actors are present and, potentially, differences in the semiotic resources available to them due to variations in hardware and software specifications and screen layout (e.g. the appearance and affordances of Skype on an Android phone as opposed to a Mac laptop or when Skype is open in full-screen mode as opposed to a minimised window). Second, *semiotic misalignment* can occur when a social actor uses or has access to a semiotic resource in his online or physical personal interaction space that cannot be seen by the other participant(s). For example, a participant may use or have access to a dictionary app/website on his or her screen or to paper notes that are not available to other participants. To bring these resources into the interaction, the social actor can verbalise/describe the resource or share it visually (e.g. by screensharing or bringing the resource within the webcam frame).

In line 4, Craig announces *here we go* but shows there is some *semiotic misalignment* between the semiotic resources to which the teacher and learners have access for the task. Indeed, he states *I've got the e:r thing here all ready to go*, demonstrating he is accessing a previously prepared disembodied resource to which learners do not have access in their interactional spaces.

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press

Similarly, in line 5, his choice of the future tense (*all you've gotta do is click on it and you'll see the page*) shows his awareness of the fact that due to *semiotic lag* (see supra), the learners do not yet have access to the URL. In line 6, as the URL becomes visible in the shared text chat space, Craig's gaze moves from the text chat towards the learners' webcam images. This gaze shift allows him to check that the link has appeared in the text chat: the learners now have the possibility to embody the link and access the online collaborative writing space. The resource in the print mode thus becomes an integral part of the pedagogical interaction, as it is embodied and disembodied for the different participants. The next higher-level action of *describing the content of the resource* (line 6) allows the learners to check that, by clicking on the URL thus embodying it, they have accessed the correct resource and ensures coordinated action.

The last sequence of Extract 2 (Figure 5), begins with a shift to the next higher-level action marked by a posture change towards the webcam and a gaze shift downwards (line 7, #10) and towards the web browser icon (#11). As Craig clicks on this icon, the disembodied online document becomes, for him, embodied (#12). In line 9, Craig announces the next task stage (*let's have a little bit of a discussion about what we have to do*) before moving to the higher-level action of *summarising previous task stage*. As he does so, he begins to type in the online document (#13). However, we notice that the learners have not yet embodied this resource. Only as Craig finishes the turn in line 9 (#14) does the avatar of one learner, Didem, appear in the top right-hand corner indicating that she is connected to the space. In line 10, Craig continues to *summarise the previous task stage* before *stating the task outcome* (write an email) and giving *key task information*. Throughout his utterance, he types these instructions into the online document. In doing so, he creates a record of the instructions to which learners can refer during the task, thereby creating frozen actions in which his instruction-giving sequence is entailed. As Craig completes this action (#15), there is joint attention on the written instructions by the learner, Didem, and Craig: the learner's cursor is now placed at the beginning of the text Craig just wrote.

Having presented a turn-by-turn microanalysis of Extracts 1 and 2, we now analyse Craig's modal configuration and density in selected representative lines from these extracts to illustrate simultaneous management of multiple higher-level actions with reference to the print mode.

4.3 RQ3: How does the teacher's modal configuration and density vary when the print mode is utilised?

Variation in Craig's modal configurations and density of his interactional space indicates his level of attention/awareness for various higher-level actions as he utilises the print mode. In this section, we examine how Craig simultaneously engages in multiple higher-level actions related to instruction-giving and evidenced in the actions performed in Extract 1, line 1 (Figure 6); Extract 1, lines 3–5 (Figure 7, Left); and Extract 2, line 10 (Figure 7, Right).

First, in Extract 1, line 1 (Figure 6), the higher-level action of announcing next task stage is foregrounded in Craig's attention/awareness. This is mainly achieved through the spoken language mode, which is absolutely necessary to perform this action, and thus has high modal intensity. Therefore, spoken language appears to have the highest hierarchical prominence to perform this instruction-giving fragment. Second, Craig's modal configuration varies throughout this spoken utterance. We observe highest modal density, achieved through modal complexity, in #2 as Craig produces the keyword *send* for this instruction-giving fragment, accompanied by the modal aggregate that comprises word stress in spoken language, iconic gesture, upwards gaze shift towards the screen (learners), posture shift towards the screen, and raised eyebrows, all of which carry medium modal intensity as they would impact the higher-level action were they discontinued. Finally, Craig's other lower-level actions, object handling (moving the cursor over the browser icon) and attempt to access the print mode (the electronic resource), carry low modal intensity for the higher-level action of *announcing next task stage*: they are not visible to learners via the webcam (due to *semiotic misalignment*) and would not change the higher-level action if discontinued.

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press

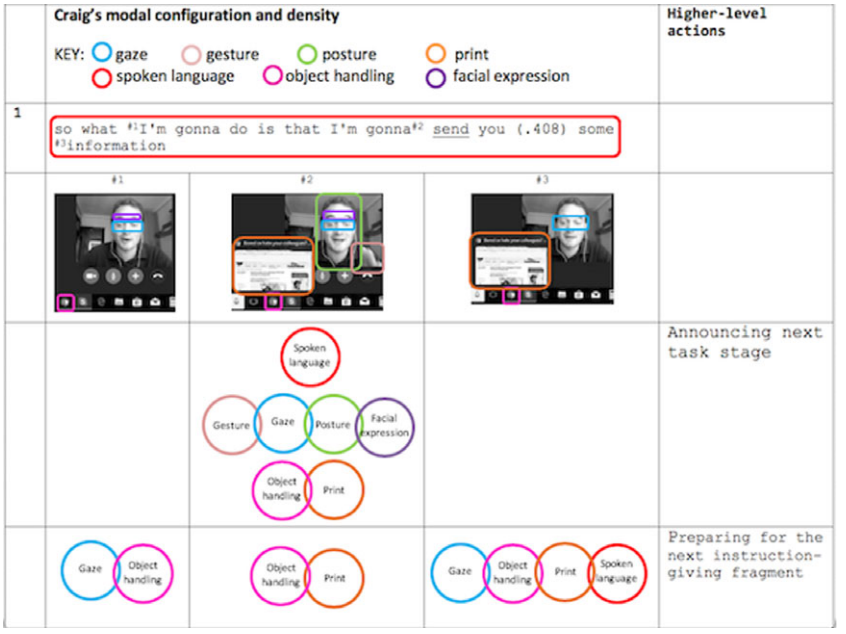


Figure 6. Extract 1, line 1. Modal configurations

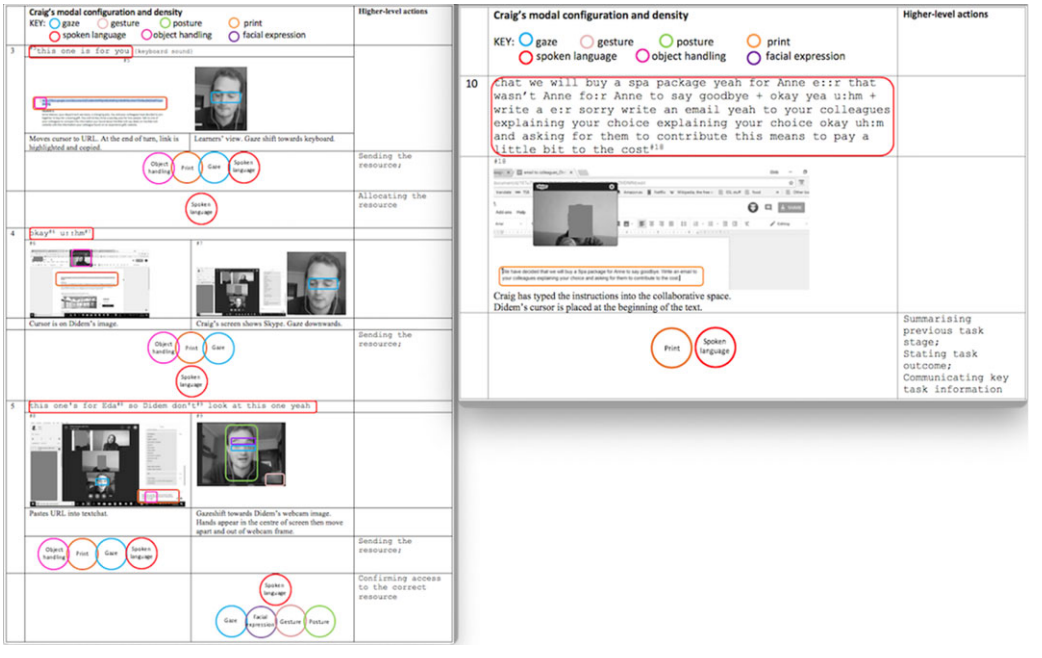


Figure 7. Left: Extract 1, lines 3–5. Modal configurations; Right: Extract 2, line 10. Modal configurations

However, Craig uses three lower-level actions (gaze, object handling, print) in line 1 to perform another simultaneous higher-level action in the background of his attention/awareness: *preparing for the next instruction-giving fragment* (i.e. *sending the resource*). This is achieved using different modal aggregates throughout #1–3. While #1 shows Craig's gaze and object handling, #2 involves

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press

object handling and print mode. However, #3 demonstrates the highest modal complexity for this higher-level action achieved through the modal aggregate gaze (towards the browser), object handling (cursor over the browser), print (webpage), and spoken language (information) as the higher-level action *sending the resource* moves to the foreground of his attention/awareness.

During Extract 1, lines 3–5 (Figure 7, Left), Craig embodies the print mode as he selects, copies, and pastes a URL for the learners to access the resource, thus performing the higher-level action of *sending the resource* (see section 4.2). For this action, although only signalled to the learners through gaze shifts, object handling and the print mode gain high modal density because the action cannot be achieved without these modes. On the other hand, spoken language (line 4) has low modal intensity as it functions as a filler, and would not change the action of *sending the resource*. In lines 3 and 4, the higher-level action *sending the resource* is thus foregrounded in Craig's attention/awareness.

Spoken language in lines 3 and 5, however, carries high modal intensity for two other higher-level actions. While Craig sends the resource, he performs the actions of *allocating the resource* and *confirming access to the correct resource*, both which cannot be accomplished without spoken language. In line 3, his modal configuration involves object handling (moving the mouse/using the keyboard to select and copy the URL), embodied print mode (the electronic resource), and downward gaze. This modal aggregate, produced alongside spoken language, has low modal density to achieve the action of *allocating the resource*. In line 5, however, the modal aggregate that accompanies the spoken language has medium modal density: its removal might impact the higher-level action of *confirming access to the correct resource*. Image #8 shows Craig's downward gaze and embodied print mode as he pastes the URL into text chat; the referent of the word *this* in the spoken language mode. Likewise, Craig's upward gaze direction towards the learner, Didem, his gesture, and his raised eyebrows (facial expression, #9) reinforce the negative imperative in the spoken language (*don't look*).

Finally, we examine Craig's modal configuration in Extract 2, line 10 (Figure 7, Right), as he embodies the print mode by typing his instructions on the shared online document while simultaneously verbalising them. During this sequence, Craig performs three higher-level actions: *summarising previous task stage*, *stating task outcome*, and *communicating key task information*. Two modes carry high modal intensity here: print and spoken language. Once Craig types the instructions, the print mode becomes disembodied and continues to entail these three higher-level actions during task completion as frozen actions. Thus, by its temporal virtue of being available for a sustained period, the print mode carries high modal intensity in performing the higher-level actions in line 10. Spoken language also has high modal intensity because although it is a think-aloud sequence that ensures bimodal production of the typed instructions (indicated by Craig's verbalisation of his spelling mistakes: *for Anne e:r that wasn't Anne fo:r Anne*); it involves repetitions and vocabulary explanations (i.e. *contribute*), discontinuation of which would significantly change the action. Consequently, modal density achieved through high modal intensity of both the print mode and spoken language foregrounds these higher-level actions in all participants' attention/awareness and becomes a joint focal point of attention.

5. Findings and discussion

To investigate an experienced online teacher's instruction-giving practices, multimodal (inter) action analysis was employed to firstly examine the site of engagement (RQ1). Our analysis demonstrated the intersection of two large-scale social practices constructed through the mediated actions within the data set: partaking in a teaching–learning activity and engaging in a research study; that is, how the data set acted as the mediational means and cultural tool within the social practice of conducting the research study. From an epistemological viewpoint, the notion of site of engagement appears to be a useful construct that could help

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press

the sociocultural researcher identify, acknowledge, reflect upon, and render explicit the theoretical lens she or he applies to the study and may contribute to documenting how research practices into online and face-to-face classrooms are fundamentally different forms of knowledge generation (cf. Stickler & Hampel, 2019).

Concerning the role of the print mode in achieving the higher-level action of instruction-giving (RQ2), within Extract 1, the print mode was employed in the higher-level actions of *sending the resource*, *allocating the resource*, *describing the content of the resource*, and *confirming access to the resource*. First, the most salient frozen actions in the print mode (i.e. preparation of the resource sheets by the researchers and appropriation of the documents by the teacher) reinforced the site of engagement. Second, the disembodied resource sheets created at a previous time by the researchers were embodied by the teacher during the lesson to share the electronic resources with learners. Third, demonstrative adjectives in the spoken language mode conveyed perception in time and space of the URLs as they are embodied and disembodied. Fourth, in the text chat screen, the URLs were represented with preview images of the documents in which the teachers' higher-level actions in this sequence became frozen and accessible to the participants for task accomplishment. Fifth, the disembodied resources that were made available in text chat also became materials waiting to become embodied again when learners read them.

Further roles were identified within Extract 2. First, the creation of frozen actions related to key task information to which learners could refer to during the activity. These frozen actions were created within the text chat (summarising the decision made during previous activity) and in the online document (giving instructions for this activity) and were available for a sustained period to facilitate task accomplishment. Second, the teacher moved back and forth between different resources in the print mode: the online document and the text chat. He embodied the URL in the previously created online document by copying it before sending it to the learners via text chat. Once sent, the link became disembodied for learners to embody it as they opened the resource. As such, in the print mode there was "constant shifting back and forth between embodied and disembodied [actions]" (Norris, 2004: 46). Third, gaze was the most salient signal for learners during the lesson for the teacher's employment of the print mode, to which they do not have access. Finally, Extract 2 illustrated that the participants interact both visually and verbally "while employing, or by employing, the mode of print" and as such "the mode of print [became] an integral part of their interaction" (Norris, 2004: 46). The latter was also elaborated upon in RQ3 (Extract 2, line 10), where we analysed the teacher's modal configuration and density during his bimodal presentation of the instructions that ensured joint attention (cf. Ricci Bitti & Garotti, 2011).

Multimodal (inter)action analysis allowed us to disassemble the modes employed in interaction in Extracts 1 and 2 through exploration of modal configuration and modal density (RQ3), which varied and indicated the relevant position of the higher-level actions within the teacher's attention/awareness as he simultaneously focused on multiple higher-level actions: *sending the resource* while *allocating the resource*, and *confirming access to the correct resource*. The modes or modal aggregates that had high modal density indicated the higher-level action upon which the teacher focused (i.e. the foregrounded action) at any given moment. Gaze direction and shifts were salient lower-level actions for signalling engagement with the print mode, whereas, for the fragment *sending the resource,* object handling combined with the print mode assumed high prominence. Although emerging CALL research is beginning to utilise multimodal (inter)action analysis as a new method (Knight, Dooly & Barberà, 2018; Satar & Wigham, 2020), modal composition of teachers' actions have yet to be revealed in such detail to inform our understanding of online language teachers' multiple foci of attention.

While Cappellini and Combe (2017) briefly describe trainee teachers' unexpected challenges in instructing learners to download a document, our turn-by-turn analysis of an experienced online teacher reveals the complexity of using electronic resources in synchronous online language teaching. Teachers need to manage such complexity to organise the semio-pedagogical activity

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press

(Guichon, 2009) and plan ahead for the higher-level actions related to instruction-giving to ensure smooth embodiment of the print mode. Teachers also need to develop necessary technological competencies to harness available semiotic resources (Hampel & Stickler, 2005) (e.g. collaborative online documents and text chat), which can be used to deliver instructions bimodally to optimise learning (Kozar, 2016; Ricci Bitti & Garotti, 2011). Successful resource management could assist teachers in creating frozen actions, which can be referred to during task accomplishment.

Given this perspective, it would be advisable for teachers to:

- set up ways of how electronic resources in the print mode (resource sheets, text chat, URLs, online collaborative writing spaces) will be used at the onset of their lessons;
- create online versions of the materials prior to the lesson;
- use clear and simple file names for the electronic documents;
- compile the materials in a single online folder or repository;
- share information around how different materials will be delivered and where they can be found;
- check learners' access to the materials before giving the instructions;
- guide the learners in the use of the electronic resources and ask them to share their screens where relevant (e.g. in small groups of one-to-one sessions);
- prepare a list of the URLs of the materials ready to be shared with the learners;
- send the resources ahead of time for the learners to study (which is a practice employed by another teacher in our corpus; see Satar & Wigham, 2020).

Having said this, it is important to caution against overly simplistic advice for best practices or pedagogical suggestions given the lack of research on the relationship between different instruction-giving practices and outcomes. In such research, the benchmark for success also requires a clear definition, whether this is related to cognitive, social, affective, or task-based outcomes. We elaborate on this further elsewhere (Satar & Wigham, 2020).

Some of these strategies and competences are demonstrated by the case described in this paper in relation to the ways in which the teacher, Craig, uses the task resources, text chat, URLs, and a shared online document. Thus, the case we present could inform training for online teachers to illustrate these semiotic and interactional competencies, but also to raise critical semiotic awareness (Guichon, 2013) pertaining to *semiotic misalignment* and *semiotic lag* (as defined in section 4.2) among the lesson participants, which can distort the shared interactional space.

6. Conclusion

Although synchronous online lessons have become increasingly popular, emerging research with regard to language teaching via videoconferencing appears to mainly investigate trainee-teachers' practices. Due to the flexibility it offers, teaching via videoconferencing allows for increased access to language learning, which could have a socio-economic impact on society. Therefore, it is important to better understand the practices and skills of online language teachers. This study bridged this gap by investigating an experienced online language teacher's use of the print mode when giving task instructions in a videoconferencing environment. The study is novel in that it adopts Norris's (2004, 2019) multimodal (inter)action analysis: a methodological framework seldom applied within CALL. By exploring how higher-level actions are constructed through lower-level actions, we were able to (1) analyse both macro and micro levels of mediated interaction situating the use of modes within social practices, and (2) reveal *semiotic misalignment* caused by computer mediation that precludes the establishment of a completely shared interactional space. The approach required us to extend Norris's (2004) definition of frozen actions being entailed in material objects in the interactional setting to include electronic objects, and her definition of the print mode to encompass electronic print. Therefore, by employing this

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press

framework, and appropriating it to our specific context, we tried to respond to the need for "researching online language learning … from new and innovative approaches, … [which] requires a conscious effort and redirection of research energies to deal with the material differences that make online language learning unique" (Stickler & Hampel, 2019: 24).

Our findings are limited to a single case study with data pieces selected from a semi-controlled corpus. Other case studies from a larger ecological corpus could further knowledge of the challenges of synchronous online environments inherent to instruction-giving and illustrate teachers' different approaches to overcome these. It would also be useful to track learners' attention/awareness through their screen-recordings while the teacher employs the print mode. This would aid understanding of the interactive, collaborative nature of instructions within a recipient design. Additional data collection methods, such as eye-tracking supported by stimulated recall interviews (e.g. Shi *et al.*, 2017) would also help better understand participants' attention and awareness. Concerning participant numbers, we incorporated two learners to allow for comparisons with previous research (Satar & Wigham, 2017), yet individual classes are perhaps more common in online teaching. To overcome this potential limitation, we have collected further data where the same task is conducted with one learner to explore whether the use of the print mode would differ in this context.

Supplementary material. To view supplementary material referred to in this article, please visit https://doi.org/10.1017/S0958344021000070

Acknowledgements. We are grateful to Newcastle University (Faculty of Humanities and Social Sciences Research Fund, 2018 Spring call) and Université Clermont Auvergne (Foreign researchers – short research visit, 2018 call) for funding towards the project titled "An Examination of Experienced Online Language Teachers' Multimodal Instruction-Giving Practices", as well as to the teacher and learner participants.

Author contributions. Both authors contributed equally to the manuscript.

Ethical statement. Ethical approval was gained from the ethics committee at Newcastle University. All participants gave informed consent prior to their participation in the study. The research project covered the teachers' normal hourly rate for each lesson completed, and a £20 e-commerce voucher compensated them for the time spent completing the post-sequence interview. Learners participated on a voluntary basis.

Pseudonyms are used. Images were edited to remove personal identifiers according to participant preferences. An earlier draft of this manuscript was sent to the participating teacher for feedback on our analyses.

References

Breen, M. P. (1987) Contemporary paradigms in syllabus design Part II. *Language Teaching*, 20(3), 157–174. https://doi.org/10.1017/S026144480000450X

Cappellini, M. & Combe, C. (2017) Analyser des compétences techno-sémio-pédagogiques d'apprentis tuteurs dans différents environnements numériques: résultats d'une étude exploratoire. *ALSIC*, 20(3). https://doi.org/10.4000/alsic.3186

Codreanu, T. & Celik, C. C. (2012) La médiation de l'interaction pédagogique sur une plateforme de visioconférence poste à poste. *ALSIC*, 15(3). https://doi.org/10.4000/alsic.2572

Cohen, C. & Wigham, C. R. (2019) A comparative study of lexical word search in an audioconferencing and a videoconferencing condition. *Computer Assisted Language Learning*, 32(4): 448–481. https://doi.org/10.1080/09588221.2018.1527359

Compton, L. K. L. (2009) Preparing language teachers to teach language online: A look at skills, roles, and responsibilities, *Computer Assisted Language Learning*, 22(1): 73–99. https://doi.org/10.1080/09588220802613831

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

Ellis, R. (2000) Task-based research and language pedagogy. *Language Teaching Research*, 4(3): 193–220. https://doi.org/10.1177/136216880000400302

Guichon, N. (2009) Training future language teachers to develop online tutors' competence through reflective analysis. *ReCALL*, 21(2): 166–185. https://doi.org/10.1017/S0958344009000214

Guichon, N. (2013) Une approche sémio-didactique de l'activité de l'enseignant de langue en ligne: réflexions méthodologiques. *Education & Didactique*, 7(1): 101–116. https://doi.org/10.4000/educationdidactique.1679

Guichon, N., Bétrancourt, M. & Prié, Y. (2012) Managing written and oral negative feedback in a synchronous online teaching situation. *Computer Assisted Language Learning*, 25(2): 181–197. https://doi.org/10.1080/09588221.2011.636054

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press

Guichon, N. & Cohen, C. (2014) The impact of the webcam on an online L2 interaction. *Canadian Modern Language Review*, 70(3): 331–354. https://doi.org/10.3138/cmlr.2102

Guichon, N. & Drissi, S. (2008) Tutorat de langue par visioconférence: comment former aux régulations pédagogiques? *Recherches en didactique des langues et des cultures – Les Cahiers de l'Acedle*, 5(1): 185–217. https://doi.org/10.4000/rdlc.6410

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

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 videoconferencing to support multimodal interaction in an online language classroom. *ReCALL*, 24(2): 116–137. https://doi.org/10.1017/S095834401200002X

Kern, R. (2014) Technology as pharmakon: The promise and perils of the internet for foreign language education. *The Modern Language Journal*, 98(1): 340–357. https://doi.org/10.1111/j.1540-4781.2014.12065.x

Kern, R. (2015) *Language, literacy, and technology*. Cambridge: Cambridge University Press.

Knight, J., Dooly, M. & Barberà, E. (2018) Multimodal meaning making: Navigational acts in online speaking tasks. *System*, 78: 65–78. https://doi.org/10.1016/j.system.2018.07.007

Kozar, O. (2016) Text chat during video/audio conferencing lessons: Scaffolding or getting in the way? *CALICO Journal*, 33(2): 231–259. https://doi.org/10.1558/cj.v33i2.26026

Lee, H., Hampel, R. & Kukulska-Hulme, A. (2019) Gesture in speaking tasks beyond the classroom: An exploration of the multimodal negotiation of meaning via Skype videoconferencing on mobile devices. *System*, 81: 26–38. https://doi.org/10.1016/j.system.2018.12.013

Markee, N. (2015) Giving and following pedagogical instructions in task-based instruction: An ethnomethodological perspective. In Jenks, C. J. & Seedhouse, P. (eds.), *International perspectives on ELT classroom interaction*. Basingstoke: Palgrave Macmillan, 110–128. https://doi.org/10.1057/9781137340733_7

McNeill, D. (1992) *Hand and mind: What gestures reveal about thought*. Chicago: The University of Chicago Press.

Mercer, R., Wegerif, N. & Dawes, L. (1999) Children's talk and the development of reasoning in the classroom. *British Educational Research Journal*, 25(1): 95–111. https://doi.org/10.1080/0141192990250107

Meskill, C. & Anthony, N. (2010) *Teaching languages online*. Bristol: Multilingual Matters.

Norris, S. (2004) *Analyzing multimodal interaction: A methodological framework*. London: Routledge.

Norris, S. (2019) *Systematically working with multimodal data: Research methods in multimodal discourse analysis*. Hoboken: John Wiley & Sons.

Norris, S. (2020) *Multimodal theory and methodology: For the analysis of (inter)action and identity*. Abingdon: Routledge.

Norris, S. & Jones, R. H. (2005) Introducing sites of engagement. In Norris, S. & Jones, R. H. (eds.), *Discourse in action: Introducing mediated discourse analysis*. Abingdon: Routledge, 139–140.

Norris, S. & Makboon, B. (2015) Objects, frozen actions, and identity: A multimodal (inter)action analysis. *Multimodal Communication*, 4(1): 43–59. https://doi.org/10.1515/mc-2015-0007

Norris, S. & Pirini, J. (2016) Communicating knowledge, getting attention, and negotiating disagreement via video conferencing technology: A multimodal analysis. *Journal of Organizational Knowledge Communication*, 3(1): 23–48. https://doi.org/10.7146/jookc.v3i1.23876

Ricci Bitti, P. E. & Garotti, P. L. (2011) Nonverbal communication and cultural differences: Issues for face-to-face communication over the internet. In Kappas, A. & Krämer, N. C. (eds.), *Face-to-face communication over the internet: Emotions in a web of culture, language and technology*. Cambridge: Cambridge University Press, 81–99.

Satar, H. M. (2013) Multimodal language learner interactions via desktop videoconferencing within a framework of social presence: Gaze. *ReCALL*, 25(1): 122–142. https://doi.org/10.1017/S0958344012000286

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, M. & Wigham, C. R. (2020) Delivering task instructions in multimodal synchronous online language teaching. *ALSIC*, 23(1). https://doi.org/10.4000/alsic.4571

Scollon, R. (1998) *Mediated discourse as social interaction: A study of news discourse*. London: Longman.

Shi, L., Stickler, U. & Lloyd, M. E. (2017) The interplay between attention, experience and skills in online language teaching. *CercleS*, 7(1): 205–238. https://doi.org/10.1515/cercles-2017-0007

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)*. http://www.lrec-conf.org/proceedings/lrec2008/pdf/208_paper.pdf

Stickler, U. & Hampel, R. (2019) Qualitative research in online language learning: What can it do? *International Journal of Computer-Assisted Language Learning and Teaching*, 9(3): 14–28. https://doi.org/10.4018/IJCALLT.2019070102

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press

Strauss, A. L. & Corbin, J. M. (1998) *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks: SAGE.

Wang, Y. (2006) Negotiation of meaning in desktop videoconferencing-supported distance language learning. *ReCALL*, 18(1): 122–145. https://doi.org/10.1017/S0958344006000814

Watson Todd, R., Chaiyasuk, I. & Tantisawetrat, N. (2008) A functional analysis of teachers' instructions. *RELC Journal*, 39(1): 25–50. https://doi.org/10.1177/0033688208091139

Wigham, C. R. & Chanier, T. (2015) Interactions between text chat and audio modalities for L2 communication and feedback in the synthetic world *Second Life*. *Computer Assisted Language Learning*, 28(3): 260–283. https://doi.org/10.1080/09588221.2013.851702

Yin, R. K. (2018) *Case study research and applications: Design and methods* (6th ed.). Los Angeles: SAGE.

About the authors

Ciara R. Wigham is a lecturer in applied linguistics and TESOL at Université Clermont Auvergne, France. Her research in the field of computer-assisted language learning examines the contribution of multimodality to pedagogical interactions, specifically within the context of virtual exchange, content and language integrated learning, and online teacher training.

Müge Satar is a lecturer in applied linguistics and TESOL at Newcastle University, UK. She is interested in technology-enhanced foreign language learning and teacher education, particularly in the concepts of social presence, meaning-making, and translanguaging in multimodal online communication, including online language teaching, virtual exchange, and blended learning.

Author ORCiD. Ciara Wigham, https://orcid.org/0000-0001-9704-1906
Author ORCiD. Müge Satar, https://orcid.org/0000-0002-2382-6740

https://doi.org/10.1017/S0958344021000070 Published online by Cambridge University Press