*ReCALL (2020), 32: 1, 25–46*
doi:10.1017/S0958344019000132

ARTICLE

# Navigating a multimodal ensemble: Learners mediating verbal and non-verbal turns in online interaction tasks

Janine Knight
Universitat Internacional de Catalunya, Spain
Universitat Oberta de Catalunya, Spain (janine@uic.com)

Melinda Dooly
Universitat Autònoma de Barcelona, Spain (melindaann.dooly@uab.cat)

Elena Barberà
Universitat Oberta de Catalunya, Spain (ebarbera@uoc.edu)

**Abstract**

Research into the multimodal aspects of language is increasingly important as communication through a screen plays a greater role in modern society than ever before (Liou, 2011). Multimodality has been explored from a number of angles relating to computer-mediated communication (CMC), such as its affordances and impact on language learners, highlighting its relevance and importance in the field of second language acquisition (SLA). Because CMC scenarios require attending to both peers and the screen, learners can be seen as positioned as "semiotic initiators and responders" (Coffin & Donohue, 2014). Increasingly, researchers are highlighting a need for a methodological "turn" to analyse this scenario from a "language" focus to a more holistic understanding of the interactions (Flewitt, 2008; Hampel & Hauck, 2006; Kress & van Leeuwen, 2001; Lamy, 2006). Along these lines, this case study explores how the action of task completion is mediated between six dyads (and individuals within the dyads) during an online peer-to-peer audioconferencing event. Drawing on notions from multimodal (inter)actional analysis (Norris, 2004, 2006) and the notion of "semiotic initiators and responders", it investigates semiotic mediation with screen-based resources through analysis of audio recordings, screenshots, log files, task simulation and reconstruction. Results highlight oral and screen-based initiations and responses that take place during task completion, which is presented as a framework.

**Keywords:** task-based synchronous communication (TB-SCMC); spoken interaction; semiotic mediation; multimodal turn-taking; screen-based resources; agency

## 1. Introduction

The many digital environments inhabited for work, play or socialisation offer ever-expanding opportunities for different types of interaction – or what van Lier (2000) has referred to as a "semiotic budget" (p. 252) for learners in online language learning programmes. In the language learning classroom, this semiotic budget pertains to the form of materials that have been designed to stimulate engagement with the target language. In online tasks, these may include the potential visual, textual and aural inputs that learners use as semiotic resources in interaction for task completion. It is therefore important to analyse how these resources form part of communication and representation in computer-mediated communication (CMC) through a multimodal lens to take into account that there is more to communication than just spoken or written language.

**Cite this article:** Knight, J., Dooly, M. & Barberà, E. (2020). Navigating a multimodal ensemble: Learners mediating verbal and non-verbal turns in online interaction tasks. *ReCALL* 32(1): 25–46. https://doi.org/10.1017/S0958344019000132

Understanding online language learning through a multimodal lens is important for a number of reasons. Visual elements through a screen, particularly multimodal texts, are gaining prominence for communication in modern society (Liou, 2011). There is a need for studying CMC task-based events through this lens because current language learning technologies incorporate increasingly more graphic, visual, textual and auditory information (Collentine, 2009) that "converge" (Herring, 2015) in different screen-based modes and digital CMC scenarios simultaneously. How the learner interacts with these convergent modes during task performance is relevant to the learning process. Thus, multimodality should be taken into account when it comes to task design (Canto, de Graaff & Jauregi, 2014; Dooly, 2018; Hampel, 2010; Hauck, 2010).

In tasks facilitated through audioconferencing, learners may be exposed to different semiotic (screen-based) resources presented as "a coherent, integrated, communicational unit" or a "multimodal ensemble" (Bezemer & Kress, 2016: 23). This scenario may require learners to be "semiotic initiators and responders" (Coffin & Donohue, 2014), not just initiating and responding to spoken language but also responding to or producing various texts, images, and so on. Learners may mediate this ensemble in many different ways – at times innovatively and differently from the initial design purpose.

Despite the availability of more channels and modes for task-based, computer-assisted language learning designs (of which CMC forms a part), there is a lack of research on the impact of multimodal communication in online language classrooms (Hampel & Stickler, 2012) and multimodality in task-based classrooms in general (Gilabert, Manchón & Vasylets, 2016). Methodological approaches to analysis are still at an exploratory stage (Rossolatos, 2015) and remain a challenge (Herring, 2015) despite a few key studies (see Flewitt, Hampel, Hauck & Lancaster, 2013; *ReCALL* special issue, September 2016). This study aims to help fill that gap by exploring how the (semiotic) mediation for task completion is carried out whereby learners may be positioned as "semiotic initiators and responders" (Coffin & Donohue, 2014).

We will first present studies that consider the underlying complexity of communication that takes place through and with technology and how technology can be seen as a participant or agent in the interaction. We then aim to create a framework for understanding the interactions taking place *with* the screen-based resources and *through* the screen, whereby orientation to potential screen-based "others" (Raudaskoski, 1999) may be more fully explored. We also want to know whether discourse descriptions of "semiotic initiations and responses" (discussed in more detail in section 2.2) are useful notions to characterise mediation taking place between humans and the screen, alongside human–human interaction. To study this, we ask, how do learners carry out actions that are mediated through and with the screen-based resources when completing an online spoken interaction task?

## 2. Literature review

### *2.1 Semiotic mediation and CMC in language learning*

The terms "semiotic resources" and "tools", introduced by Vygotsky (1981) and highlighted by van Lier (2000) and Lantolf (2000) within a sociocultural perspective of language learning, were coined with a view to understanding how language learning is mediated through different available resources that students encounter during online tasks. During task completion, semiotic resources can mediate goal-directed actions. Mediated action involves agents and their cultural tools – both are mediators of the action (Wertsch, 1998). In this sense, Wertsch (1994) aimed to underscore the "recognition that humans play an active role in using and transforming cultural tools and their associated meaning systems" (p. 204) while "at the same time, however, any instance of mediated action involves a reiterative dimension" (p. 206). It can be argued then that language acquisition occurs through a dynamic process of the self, interacting with cultural tools; this "mediation" includes the use of language as well as other tools as a social resource, which

https://doi.org/10.1017/S0958344019000132 Published online by Cambridge University Press

becomes internalised by the learner. This cognitive development occurs moment by moment in social interaction (Lantolf, 2000), and therefore a microanalysis of discourse in its sequential context can allow the researcher to examine this process in motion (Lantolf, 2000).

An appreciation of how cultural tools or mediational means are involved in human action (including learning) forces us to go beyond the individual agent when trying to understand the forces that shape such action (Wertsch, 1998). Therefore, attending to "the material *stuff*" (Kress, 2003: 32, original emphasis echoed by Lamy, 2006), in this case the materiality of the screen, suggests that any analysis of mediated action for the completion of a spoken interaction task should consider multiple ways of understanding the action.

In online CMC scenarios, these complex situations involve mediation through human–human oral interaction as well as mediation that takes place as a result of human–computer interaction. In order to understand such situations, it is necessary to expand the analysis to include non-embodied screen-based semiotic resources. By non-embodied modes we mean screen-based resources that may not form part of the direct two-way *communication* but rather are *represented* from computer to learner. To exemplify, in CMC facilitated by audioconferencing, audio is a mode emanating from the oral utterances of their partner *through* the channel, but it is also possible for audio to be a mode represented by the screen *with* which learners can also mediate their learning (e.g. the sound from video clips). Learners can be both "semiotic initiators" and "semiotic responders" (Coffin & Donohue, 2014): initiating and responding orally with their peer through oral turns and turn-taking as well as to various screen-based resources.

There are increasingly more CMC studies for spoken interaction that highlight the screen-based modes and interface. Hampel and Stickler (2012) identified communication modes in a videoconferencing event as linguistic (spoken and written) alongside visual such as icons (vote buttons, yes/no/?, emoticons), still and moving images, display/scrolling of text and gestures. Lamy's (2006) study on turn-taking and face-saving using an audiographic tool identified natural language (written and spoken) as well as visual resources (icons, images, colours and shapes). Vetter and Chanier's (2006) study on how language learners used multimodal tools to make spoken interactions highlighted text, speech, and graphics for communication as well as the interplay of modes (text and spoken language). Knight and Barberà's (2016) study of peer-to-peer spoken interaction tasks using an audioconferencing tool found that learners were multitasking as they interacted with language (text), image (photo), icon (pop-up) and navigational resources and that the different screen-based resources appeared to relate to different learner purposes. Balaman and Sert's (2017) study on two different task types in two different settings (face-to-face and online using audioconferencing) also highlighted screen-based modes – in their case, the video clips on the screen whereby learners could type answers, click on answer buttons and receive correct answers whilst conversing with their partner. They also noted long silences that potentially pointed to the ongoing orientations to the task interface and screen-based resources.

### 2.2 Semiotic initiation and response and multimodal turn-taking

"Semiotic initiation" and "response" (Coffin & Donohue, 2014) in CMC tasks in the field of SLA has largely been studied through the analysis of verbal turn-taking, often using conversational analysis (CA). Whereas CA was originally focused on producing a purely verbal outline of the turn-taking "system" (sometimes referred to as the "speech exchange system"), as proposed by Sacks, Schegloff and Jefferson (1974), there has been a growing interest in the multimodal dynamics of the turn-taking process in the field of linguistics, including oral turn-taking in SLA (see overview by Jenks, 2014). More recent studies (some beyond the field of SLA) have expanded oral turn-taking with other modes such as gaze behaviour (e.g. Oben & Brône, 2015) and gestures and have taken body positioning and spatiality into account (e.g. Mondada, 2007, 2013).

https://doi.org/10.1017/S0958344019000132 Published online by Cambridge University Press

Some analysts, including many who use a CA approach, both in online language learning and non-language learning environments, found that turn-taking can take place transmodally – across modes (Lamy, 2006, in audiographic conferencing; Helm & Dooly, 2017; Liddicoat, 2010, in videoconferencing). Both Lamy (2006) and Helm and Dooly (2017) highlight the use of a hybrid, mixed-mode interaction (with text and speech), often at times resulting in time lags and overlapping of turns. Key to this understanding of online, multimodal turn-taking is the way in which speakers in each oral turn demonstrated to one another their own understanding of the previous speakers' oral turn and that these aspects of turns and turn-taking were context sensitive to both task type and task setting (cf. Balaman & Sert, 2017). These authors show that learners coordinate turn-taking through their mutual alignment to alignments to screen-based resources, online oral interaction, and other features of the interface.

Many CA analysts maintain that CMC conversations can involve a multimodal accomplishment of openings, interruptions and closings as oral turns carried out in various mediums (Tudini, 2014). Liddicoat (2010) highlighted a further complexity: it can be argued that turn-taking can take place between humans and screen-based resources in addition to between humans. In his study, Liddicoat (2010) found that the initiator in the beginning of an online conversation must first capture the attention of a non-present co-participant through technology. This is achieved by a message via the computer (screen), namely "Andrew wants to have a video conversation". This message was neither spoken nor written by Andrew but is initiated by him through the technology. The response of his partner was either a choice to press "respond" or "refuse". This resembled summons-answers sequences where the verbal equivalent may be "hey" or naming and the technological equivalent would be the ringing of a telephone (Liddicoat, 2010). The online nature of the participants' interaction was considered a relevant constituent part of the interactions (turns) and not just a facilitative one (Liddicoat, 2010).

Beyond CA studies, turn-taking through computers has been extensively studied in the field of human–computer interaction (HCI) and CMC; however, a thorough review of these studies is beyond the scope of this paper. Pertinent (non-SLA) studies have highlighted that screen-based resources can be "active agents" that send reminders (Dourish, et al., 1993), act as agents in conversation characterised by "presentation" and "acceptance" phases (Clark & Brennan, 1991) and can be used as conversational resources in the accomplishment of physical tasks (Kraut, Fussell & Siegel, 2003).

More recently, Benson (2015), taking a digital discourse perspective, investigated physical turn-taking with the interface of YouTube and employed the notion of "orientation" from CA to understand turns. He operationalised "responses" on a YouTube page as video responses, "like"/"dislike" icons or written comments. Turn "initiation" included uploading a video – highlighting the visual, textual and physical modes involved. What these HCI studies highlight is that screen-based resources can appear to act out turns (as initiators or responders). Furthermore, the intentional physical "moves" of human participants (e.g. typing a response or clicking "like"/"dislike" icons) in relation to the screen-based resources could also be considered part of initiation/response sequences.

Taking this into account, the screen-based resources can be understood through the notion of potential "others" that may also act as and/or be orientated to (Raudaskoski, 1999). The study of "others" has generally been approached as being (1) text or discourse, (2) a social entity or agent or (3) a sign (Raudaskoski, 1999). In "encounters" with screen-based resources, as opposed to "conversations", humans can be positioned as "others" and screen-based resources can also act as "others" (Raudaskoski, 1999). However, because the semiotic resources are not "fully fledged communicative partner[s] . . . ", "the sense making has to be constructed one-sidedly, rather than coconstructed, making the human participant solely responsible for the emerging meaning" (Raudaskoski, 1999: 22–23).

## 3. Methodology

### 3.1 Context and participants

The participants were students in an English-as-a-foreign-language class in an online degree programme. The learners were B1.1 level on the CEFR (upper intermediate) group. There were 12 adult students: two male and 10 female, 26–55 years old, engaged in a virtual synchronous peer-to-peer oral role-play task. Students were bilingual (Catalan and Spanish) with English as an additional language. Participant names have been changed.

Participants were presented with the Tandem audioconferencing tool (http://www.speakapps.eu/#tandem), a content management application that distributed the task materials and provided prompts to continue the task in immediate response to the participants' actions. The 12 students formed six dyads. Data sources included approximately 97 minutes of peer-to-peer recorded oral conversations of which approximately 53 minutes were transcribed roughly before learners changed roles and repeated the tasks. Screenshots of the task (taken by the researcher) and Tandem tool logs (which indicated the date, time, number of entries and duration the tool was open for use) were also used in the analysis.

The task was a role-play task (divided into subtasks) in which learners took turns being the interviewer and interviewee. The first task required that one learner ask questions to their peer in their roles. The second subtask required the interviewer to describe two jobs that included details visible only to the interviewer. After listening to the jobs, the interviewee needed to indicate their preferred job. After the first two subtasks were completed (Task 1 and 2 in the screenshots) the subtasks were repeated but the peers changed roles (Tasks 3 and 4 in the screenshots). There was a timer that indicated how many minutes and seconds were left to complete the task. This was followed by a pop-up that appeared when the predetermined time was up (four minutes for Task 1 and seven minutes for Task 2). Only the first two tasks were analysed, as this was deemed sufficient to answer the main driving question. The analysis stopped at the point when learners swapped roles and began to repeat the two subtasks. This meant that one student was analysed as interviewer. The different screen-based resources included textual/visual task instructions, textual prompts to help learners form oral questions, textual instructions about learner roles and technological aspects of the task, navigational resources to move one subtask on to the next, and a pop-up indicating the time was up among other non-task-related resources around the periphery of the screen.

### 3.2 Approach and analysis

A qualitative case study approach was employed incorporating a purposive sampling procedure in order to select dyads that had appeared to follow task design and others that appeared to diverge from it. Data sources were triangulated with expert opinions and checked with colleagues regarding the tasks given and task conditions. The approach involved three main interrelated foci of analysis and phases.

All of the foci took "mediated action" as the unit of analysis (Wertsch, 1998). We operationalised mediated action by drawing on notions from multimodal (inter)actional analysis developed by Norris (2004) and notions from CA (Sacks et al., 1974), in particular the orchestration of turn-taking. However, this study did not rigorously follow the protocols of CA transcription as the main focus and analysis was not on oral turn-taking but rather on the non-oral sequentiality of initiations and responses (potentially with the screen). However, the principles of "relevance" and "orientation" from CA, which Benson (2015) used in his identification of participants' physical turns with the screen (e.g. clicking on, uploading), were drawn upon. We categorised learners' physical turns, which can be initiations or responses with the screen (e.g. when learners indicated that they had clicked on a screen-based resource). In addition, we operationalised

https://doi.org/10.1017/S0958344019000132 Published online by Cambridge University Press

**Table 1.** Phases of data analysis

| | Activity | Aim |
|---|---|---|
| Phase 1 | Collecting explicit mentions of resources by the participants | Identify the trajectories of screen-based resource use from a learner perspective |
| Phase 2 | Reconstructing possible task sequence | Identify the trajectories of screen-based resource use from a researcher perspective |
| Phase 3 | Mapping of explicit mentions to reconstructed task sequence<br>Comparing and contrasting different cases to specific screen-based initiations or responses | Gain an overview of the trajectories of screen-based resource use (both perspectives)<br>Gain further insight into the mediation process |

screen-based turns or initiations as evidenced when learners responded to them orally, physically, visually or using a combination. The three analytical phases are depicted in Table 1.

In phase 1 the researcher noted the resource use from the learners' perspective: noting the resources that learners explicitly mentioned orally in the audio recordings (and therefore those that learners deemed as "relevant", according to CA). These were collected in a table and labelled "L" for learner perspective. Notes were made iteratively, dyad after dyad, which were then transformed into a comparative table to facilitate a cross-case analysis. This process resulted in a chronological overview of the trajectories of screen-based resource use made explicitly relevant by the pairs in their recorded talk, which was later expanded with more examples (next step explained as follows). The concept of "relevance" reflected the notion that "modes do not exist without social actors utilizing them in some way" (Jewitt, Bezemer & O'Halloran, 2016: 115).

Phase 2 utilised a second lens for analysis, namely the researcher's perspective. The main researcher reconstructed the learner's "steps" (who had the interviewer's role) as if they were following the task design. This was carried out through task simulation with a colleague. Screenshots were taken of the process and reconstructed in a document to see the sequence of the task process (see Appendix A). However, the screenshots did not use recordings of the participants' screens, but rather the screen of the researcher during task simulation.

Phase 3 involved listening to the recordings with the screenshots in hand. More specifically, it observed the learners' interaction with the screen by (a) listening to the learner in combination with (b) the researchers' screenshots that simulated the task process. By doing so, the researcher could follow if and how the learner responded to the screen-based resources. This was repeated many times in order to identify instances of "semiotic initiations" and "responses" (Coffin & Donohue, 2014) (that also encompassed oral turn-taking) from both learner and screen. We focused on learners' oral turns as responses to previous oral turns and learners' potential response/initiations to the screen-based resources. We identified (a) the topic of the turn, (b) the learner's understanding of the previous turn/initiation and (c) if and how learners "react to the messages" (Norris, 2006: 4). The identification of resources that initiated a turn (such as a pop-up) or were responded to during this analysis was transposed onto the initial table and labelled "R" for researcher perspective. This phase also provided further insight into the mediation process through repeated simulation of the task for each case with the screenshots in hand while listening to and comparing different cases. Researcher notes were made and other tables were constructed in order to compare various similarities and differences between the cases and individual learners' behaviour. This yielded a focus on both the jointly constructed mediated action of the dyad and the individually mediated action of the learner/interviewer with the screen.

Apart from the processes previously described, a detailed analysis of the screen and subsequent labelling process of the screen-based resources and what appeared on the screen in between (see Appendix B) provided further understanding of available resources for the participants. These were then categorised in accordance with Lamy (2006), who identified natural language in its

https://doi.org/10.1017/S0958344019000132 Published online by Cambridge University Press

written and spoken forms, as well as visual resources such as icons and images. In our categorisation we added "navigational" to "textual" and "visual" as well as the terms "static" and "dynamic" to indicate whether screen-based resources were moving or not, with the latter two labels taken from Herring's (2015) rubric of multimodal CMC. This was useful for a detailed view of a key term in our analysis – "resource" (e.g. a visual timer, pop-up on the screen) – and how we could categorise them. Because resources were made up of a number of modes (e.g. textual and visual), we labelled the resources according to the sequence/hierarchy of the modes that were important to each resource's designed purpose. So, for example, a pop-up that appeared suddenly needed to be seen first, then read to know what to do with it, and then "closed" with a physical click. This was labelled as a visual/textual/navigational resource. Textual information pertaining to the task was labelled as textual/visual, whereas banners designed to signal task sequence were labelled visual/textual. However, we recognise that this was problematic depending on what mode learners were attending to at any given moment. Appendix C illustrates the analysis of the interface pages and the labelling according to this logic.

Finally, in order to bring together our understanding of mediated action (Wertsch, 1998; discussed previously) with a multimodal perspective, we propose a complementary use of multimodal (inter)actional analysis. This analytical approach identifies two levels of social (mediated) action: higher-level and lower-level actions, each of which deals with a different level of interaction. *Higher-level action* is used to refer to large-scale actions, such as a meeting, and is made up of a "multiplicity of chained lower-level actions" (Norris, 2004: 11[1]). *Lower-level action* is used to refer to smaller-scale actions, for example, gestures or gaze shifts that become chains of lower-level interactions (Norris, 2004). The lower-level actions support the achievement of the higher-level action. Higher-level mediated actions are those actions that social actors usually intend to perform and/or are aware of and/or pay attention to (Norris, 2016).

Multimodal (inter)actional analysis also deploys the notion of levels of simultaneous awareness/attention, namely *foreground, mid-ground* and *background*. This is a continuum that facilitates the visual representation of various levels of attention that an individual is simultaneously engaged in (Norris, 2016) while completing an action. A person can be engaged in various actions at a particular point in time (e.g. engaging in a research project, Skyping with family members, interacting with a girlfriend) (Norris, 2016); the variability in attention/awareness in one of these activities could be described along this continuum.

## 4. Results and discussion

Results suggested that learners mediated the action of task completion through the use of various navigational, textual and visual screen-based resources as well as with spoken language and that these resources formed part of initiation/response sequences for task completion. The three main identified strands of the turn-taking were (1) learner responses relating to pedagogical task instructions, (2) oral initiations and responses and (3) initiations/responses relating to navigational resources. In addition, a number of results pertained to the nature of mediating with the screen-based resources that appeared to highlight the various levels of attention that an individual is simultaneously engaged in (Norris, 2016). First, we present the basic semiotic initiation and responses identified in the data. Then, we show how participants mediated with the different screen-based resources through the perspective of Norris's (2004) notions of *foregrounding* and *backgrounding* from multimodal (inter)actional analysis. This is followed by a framework that outlines and discusses the initiation/response sequences identified in the analysis

[1] Norris also makes reference to "frozen actions" that are entailed in material objects after the action has taken place, such as the layout of a room. However, because we could not observe the physical space surrounding the learners' computers, this current study did not apply the notion.

https://doi.org/10.1017/S0958344019000132 Published online by Cambridge University Press

and how these sequences can also be understood using notions of higher-level and lower-level actions, equally inspired by multimodal (inter)actional analysis.

**Oral turn-taking.** Regarding spoken language, examples 1 and 2 show how learners typically used verbal turn-taking to mediate the process of pedagogical task completion.

**Example 1:** Question/answer (case 2)

L: Erm Well, (paper rustling) Thank you in advance for your time. I make you some questions about your Curriculum Vitae. First: what your complete name, please?
P: My name is Paulo Martinez.
L: Um What are your academic experience?
P: I have a degree of Psychology

Some of the oral turn-taking revealed learners' orientations to the screen-based resources. Case 3, as shown in example 2, showed how that when technological problems arose during the role-play interview and learners had to "close" or minimize a resource so they could continue, oral turn-taking through questions and answers was maintained. Learners orient to the interface ("close" icon), as found by Balaman and Sert (2017), and use the text "close" as a conversational resource in the accomplishment of the physical task, as found by Kraut et al. (2003).

**Example 2:** Question/answer (case 3)

A: Okay . . . well ok. now . . . close, no, I suppose . . . ?
F: I suppose too, I close it . . .

**Semiotic initiations and responses.** In addition to oral turn-taking, other initiation and response sequences were identified pertaining to other modes and resources – more specifically, learner responses to pedagogical task instructions and prompts as well as initiation and responses relating to navigational resources. Table 2 shows if and when learners responded to the various resources. This comprises resources that were explicitly mentioned by learners in the audio recordings, labelled in the table as "L". In addition, the resources identified as initiating, responding or being responded to in some way by the researcher afterwards were labelled as "R". This labelling was based on the task simulation and listening to audio recordings. The results from this researcher perspective were produced through a process of notetaking and tracking each case, which was finally transposed to the table. The sense-making that occurred between the screen and interviewer can be understood as being constructed one-sidedly by students while co-constructing oral turns with their partner. In addition to indicating which resources each case mentions or responds to, the table offers a general representation of learners' trajectories of use of various screen-based resources.

### 4.1 Learner responses to pedagogical task instructions and prompts

With respect to learner responses to pedagogical task instructions and prompts, in task 1 all cases responded to two pedagogical task instructions. These were namely *Use it to create questions and find out some important information about Student B* (with the "it" referring to the sample information) and *Ask your partner at least five questions*, as shown in Figure 1. We consider the pedagogical task instructions to be a screen-based textual/visual "initiation" that requires a visual/oral "response" from learners because it is a "request" to act. Learners' responses were in the form of a series of oral turns in their respective roles. Figure 1 highlights the textual/visual resources that learners respond to orally.

https://doi.org/10.1017/S0958344019000132 Published online by Cambridge University Press

**Table 2.** Screen-based resources that were and were not made verbally explicit by cases

| | (1)<br>Waiting for confirmation<br>T1 | (2)<br>Information about roles – Student A/B<br>T1 | (2)<br>Ask the minimum of 5 questions[a]<br>T1 | (2)<br>Textual sample questions; e.g. "Where do you live?"<br>T1 | (2)<br>Create own questions + symbol?[a]<br>T1 | (2)<br>Text from sample candidate<br>T1 | (3)<br>Pop-up "Time up!"/ "close" button[a] | (4)<br>Next task[a] | (2)<br>Uses text describing two jobs<br>T2 | (3)<br>Pop-up "Time up!" /"close" button[a] | Interface page 2 "Solutions page" | (5)<br>Timer | (4)<br>Next task[a] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | L | R | R | R | R | R | L | L | R | L | L | L | L |
| Case 1 | X | X | √ Inter | √ Inter | √ Inter | √ Inter | X | √ Inter | √ Inter | √ Intee | X | X | √<br>Intee |
| Case 2 | X | X | √ Inter | X | √ Inter | √ Inter | X | X | √ Inter | X | X | X | X |
| Case 3 | √ (before start T1)<br>Inter | √ Inter<br>Intee | √ Inter | √ Inter | √ Inter | √ Inter | √ Inter<br>Intee | √ Inter | √[b] Inter | √ Intee | √ Inter | X | X |
| Case 4 | X | √ T1<br>Inter<br>Intee | √ Inter | √ Inter | √ Inter | √Inter | X | X | √ Inter | X | X | X | X |
| Case 5 | X | X | √Inter | √ Inter | √ Inter | √ Inter | X | X | √ Inter | X | X | X | X |
| Case 6 | X | X | √ Inter | √ Inter | √ Inter | √ Inter | X | X | √ Inter | X | X | √[c] Inter | X |

*Note*. Numbers correspond to type of mode(s) that the resource pertains to: (1) Visual/textual (dynamic) (e.g. pop-up requiring no response), (2) Textual/visual (static) (e.g. information and instructions), (3) Visual/textual/navigational (dynamic) (e.g. pop-up requiring response), (4) Visual/textual/navigational (static) (e.g. navigational resource), (5) Visual (dynamic). L = learner perspective: resource was made relevant by learner; R = researcher perspective: resource was identified by researcher; Inter = interviewer orally, explicitly mentions the resource; Intee = interviewee orally, explicitly mentions the resource.

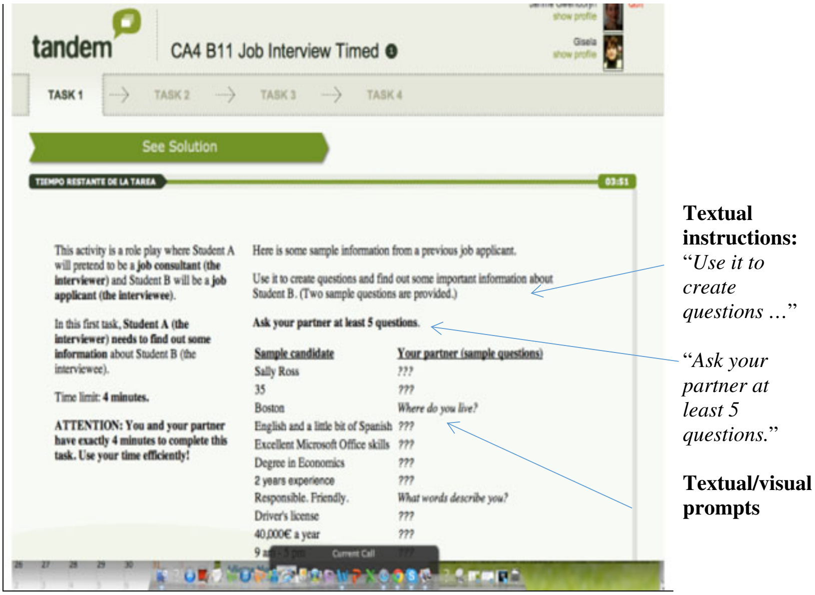
[a] The resource is considered to be a turn in the form of a request or invite that learners can respond to or accept in order to complete the turn.

[b] Case 3 refers to "the pdf" after tool does not work.

[c] Case 6 refers to the time as "when the time out we change roles". There is no other mention of navigational resources in the audio recording. This is interpreted as learners timing themselves but not necessarily with the screen-based timer.

**Table 3.** Number of questions asked by interviewer in Task 1

| Case | Number of questions asked by interviewer |
|---|---|
| 1 | 12 |
| 2 | 9 |
| 3 | 9 |
| 4 | 8 |
| 5 | 4 |
| 6 | 9 |



This activity is a role play where Student A will pretend to be a **job consultant (the interviewer)** and Student B will be a **job applicant (the interviewee)**.

In this first task, **Student A (the interviewer) needs to find out some information** about Student B (the interviewee).

Time limit: **4 minutes.**

**ATTENTION: You and your partner have exactly 4 minutes to complete this task. Use your time efficiently!**

Here is some sample information from a previous job applicant.

Use it to create questions and find out some important information about Student B. (Two sample questions are provided.)

**Ask your partner at least 5 questions.**

| Sample candidate | Your partner (sample questions) |
|---|---|
| Sally Ross | ??? |
| 35 | ??? |
| Boston | *Where do you live?* |
| English and a little bit of Spanish | ??? |
| Excellent Microsoft Office skills | ??? |
| Degree in Economics | ??? |
| 2 years experience | ??? |
| Responsible. Friendly. | *What words describe you?* |
| Driver's license | ??? |
| 40,000€ a year | ??? |
| 9 a[illegible] - 5 pm | ??? |

**Textual instructions:** *"Use it to create questions …"*

*"Ask your partner at least 5 questions."*

**Textual/visual prompts**

**Figure 1.** Textual/visual resources

All interviewers asked at least five questions (which one of the instructions explicitly requested), except for case 5 as shown in Table 3. Notably, many interviewers asked approximately 10 questions. The reason for this may be that the textual instruction to ask at least five questions was followed underneath by 11 textual/visual prompts for the interviewer (sample candidate and sample questions), as shown in Figure 1. Therefore, the number of learner turns as a response to the instruction may have been shaped more by the number of available textual/visual prompts on screen over the textual instruction/request to create five questions. Rather than having "reinforcing roles" (Kress & van Leeuwen, 2001) in the interaction between modes, the 11 prompts may have overridden the textual instruction (request) with respect to the number of questions that were asked.

### 4.2 Initiation and responses relating to navigational resources

In addition to the initiation and response sequence pattern identified previously, initiations and responses were also identified in relation to the navigational resources. These navigational resources were identified as inviting and being responded to by human physical responses to

https://doi.org/10.1017/S0958344019000132 Published online by Cambridge University Press

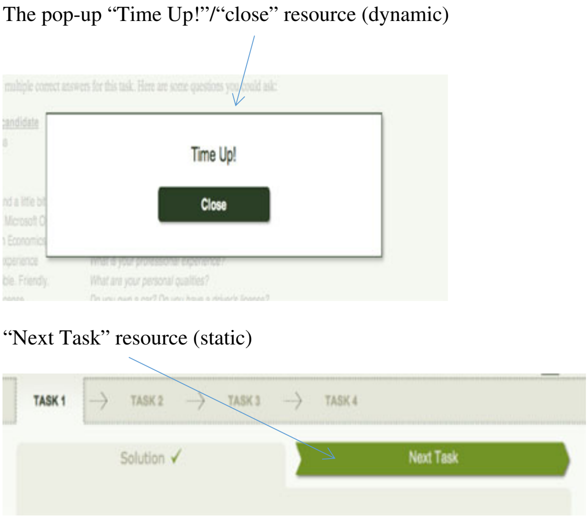

The pop-up "Time Up!"/"close" resource (dynamic)
Time Up!
Close
"Next Task" resource (static)
TASK 1
TASK 2
TASK 3
TASK 4
Solution ✓
Next Task

**Figure 2.** Navigational resources as invites for acceptance

them, which were sometimes accompanied by speech. These oral mentions were labelled as "L" in Table 2.

Cases 1 and 3 verbally mentioned the pop-up "Time Up!"/"close" button and the "next task" resource, as shown in Figure 2. However, this mention did not occur both times these resources (would have) appeared on the screen. This observation leads to the identification of mediation of turn completion through the use of navigational resources (and with) navigational resources as initiators or responders. Figure 2 highlights a navigational resource that "invites" the learner to "close", which learners then need to physically click to "accept". The turn completion is co-constructed with the navigational resource that can be understood as a turn-taker in the completion of a turn.

Both the pop-up "Time Up!"/"close" button and "Next Task" button (static) require the learners to navigate in addition to negotiating when learners navigate together. We consider the "close" button on the Time-Up! pop-up and the "Next Task" button to be "invites" that learners have to "accept" by clicking them physically. The completed action is performed individually and multimodally: visually/textually from the computer (with the block of colour and then text; "close" and "Next Task") and then relating to the students' body as they physically click these. This highlights the "mode of touch" (Bezemer & Kress, 2014: 85). Both cases 1 and 3 accompany this process with speech. This suggests that an initiation/response sequence may be constructed between learners and screen-based resources as "agents" (Clark & Brennan, 1991; Dourish, et al., 1993).

The pop-up "Time Up!"/"close" button in particular, which would have occurred twice during the two tasks, would have surprised learners. However, cases 2, 4, 5 and 6 do not make any explicit reference or utterance of surprise. From this, we deduced that their trajectory with the navigational resources was somewhat different in comparison to cases 1 and 3. A possible explanation is that because the participants were already familiar with the tool and its pop-up features, their use had become normalised in their practice. Alternatively, they had already familiarised themselves with the resources prior to starting the task. We propose that the latter was the case because the tool is not used as a common practice throughout the course so the students are not familiar with it.

The finding that the trajectories of 2, 4, 5 and 6 were different to that of cases 1 and 3 was complemented with other sources and methods including the Tandem logs, a focus of time spent

https://doi.org/10.1017/S0958344019000132 Published online by Cambridge University Press

**Table 4.** Semiotic initiations and responses identified during the analysis

| Initiation | Response |
|---|---|
| Oral turn (learner) | Oral turn (learner) |
| "Confirm to start" (computer pop-up) | Click "start" (learner) |
| "Ask your partner at least 5 questions" (screen-based text instruction as a request to act) | A series of oral turns as questions (learner) |
| Oral questions constructed by learner using textual information from the screen and "?" symbol | Oral turns (partner response to questions) |
| Time Up! (computer pop-up) | Click "close" navigational resource (learner) |
| "Describe jobs" (screen-based text instruction as a request to act) | Oral turns (constructed by learner with text as a conversational resource) |

on task and the number of seconds before starting a new oral turn with their partner. The Tandem logs, collected from approximately 50% of the cases, indicated that the learners not only used navigational resources to respond to screen-based invites but also exited and re-entered the tool. We conceptualise the learners' response to these navigational "requests" as lower-level actions, which, when combined with other completed actions, lead to the higher-level action of managing the Tandem tool. The click of the button ("close" or "Next Task") can be understood as "accept" going forward as a completed (computer–human) turn for cases 1 and 3. However, we propose that the other cases, by entering and exiting the tool, navigating back and forth, override the initial meaning of "accept", potentially reassigning a different meaning to this resource.

The amount of time that learners spend completing the two tasks confirmed the different trajectories. Cases 1 and 3 spent the minimum amount of time required for task completion according to task design with the pop-up and timers (11 minutes), whereas cases 2, 4, 5 and 6 spent less than the required time (see Appendix D). Cases 2, 4, 5 and 6 appeared to have self-regulated their time on task, which is arguably easier to do without the presence of a timer and "interruption" of a pop-up.

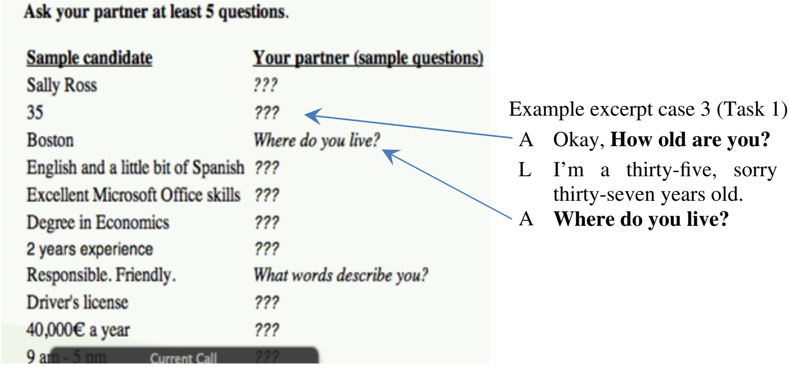
These results suggest that while some learners were prepared to carry out the mediated action of task completion using the tool and resources according to task design (i.e. spontaneously), others were not. Some dyads appeared to prefer to mediate through the use of screen-based resources or the tool before carrying out the recording (whether for linguistic reasons or reasons pertaining to the tool). This meant that they could control the task conditions (controlling their own time and navigational moves) including if they were going to pay attention to the pop-up "Time Up!"/"close" button or not. To summarise, Table 4 outlines the semiotic initiations and responses identified during the analysis.

### 4.3 Insights into the mediation process: Foregrounding and backgrounding

Insights about foregrounding and backgrounding (Norris, 2004) pertaining to the mediation process were also noted. These insights relate to learners' interactions with the different screen-based resources. They include the use of screen-based text as a conversational resource as part of their oral turns with their partner, silences as evidence of orientations to the screen, and different navigational trajectories as a way of mediating differently with the audioconferencing tool.

During the tasks, the screen-based text and the "?" sign was noted as being used by learners as conversational resources for creating oral turns, both responses and initiations. In Task 1, learners responded to the textual sample information by creating questions as oral turns for their partner (as requested). This can be seen in Figure 4 where the interviewer is beginning to ask questions. The arrows in Figure 3 and the bold text in the transcripts in Figure 4 indicate spoken words that

https://doi.org/10.1017/S0958344019000132 Published online by Cambridge University Press



Ask your partner at least 5 questions.

| Sample candidate | Your partner (sample questions) |
|---|---|
| Sally Ross | ??? |
| 35 | ??? |
| Boston | *Where do you live?* |
| English and a little bit of Spanish | ??? |
| Excellent Microsoft Office skills | ??? |
| Degree in Economics | ??? |
| 2 years experience | ??? |
| Responsible. Friendly. | *What words describe you?* |
| Driver's license | ??? |
| 40,000€ a year | ??? |
| 9 am - 5 pm | ??? |

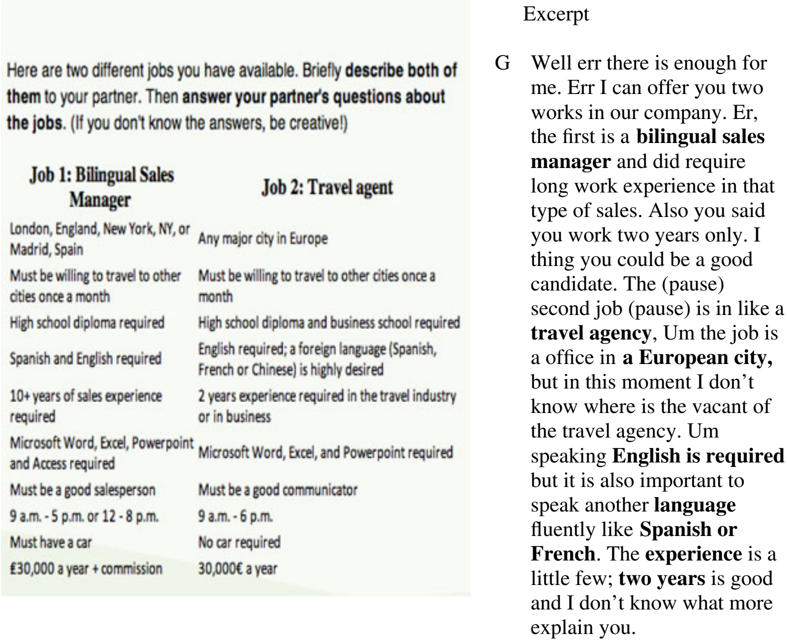
Example excerpt case 3 (Task 1)

A Okay, **How old are you?**

L I'm a thirty-five, sorry thirty-seven years old.

A **Where do you live?**

Figure 3. Textual prompts as a conversational resource in the creation of oral turns



Here are two different jobs you have available. Briefly **describe both of them** to your partner. Then **answer your partner's questions about the jobs**. (If you don't know the answers, be creative!)

| Job 1: Bilingual Sales Manager | Job 2: Travel agent |
|---|---|
| London, England, New York, NY, or Madrid, Spain | Any major city in Europe |
| Must be willing to travel to other cities once a month | Must be willing to travel to other cities once a month |
| High school diploma required | High school diploma and business school required |
| Spanish and English required | English required; a foreign language (Spanish, French or Chinese) is highly desired |
| 10+ years of sales experience required | 2 years experience required in the travel industry or in business |
| Microsoft Word, Excel, Powerpoint and Access required | Microsoft Word, Excel, and Powerpoint required |
| Must be a good salesperson | Must be a good communicator |
| 9 a.m. - 5 p.m. or 12 - 8 p.m. | 9 a.m. - 6 p.m. |
| Must have a car | No car required |
| £30,000 a year + commission | 30,000€ a year |

Excerpt

G Well err there is enough for me. Err I can offer you two works in our company. Er, the first is a **bilingual sales manager** and did require long work experience in that type of sales. Also you said you work two years only. I thing you could be a good candidate. The (pause) second job (pause) is in like a **travel agency**, Um the job is a office in **a European city,** but in this moment I don't know where is the vacant of the travel agency. Um speaking **English is required** but it is also important to speak another **language** fluently like **Spanish or French**. The **experience** is a little few; **two years** is good and I don't know what more explain you.

Figure 4. Text as a conversational resource in the creation of oral turns

were lexically the same or similar to the screen-based text or as emerging from learners' mediation with the "?" These results confirm the use of "others" as text (Raudaskoski, 1999) and also formed part of learners' response to the textual task instruction to ask their partner questions. Learners can be said to be foregrounding or backgrounding (Norris, 2004) the text-based resources through choices that they make regarding what text they use and do not use in their oral turns.

Whereas in Task 1 learners do not generally go beyond the scope of the sample information as a conversational resource, their behaviour is different in Task 2. In Task 2, learners are requested to respond to the textual instruction relating to two jobs presented: *Briefly describe them to your partner. Then answer your partner's questions about the jobs (if you don't know, be creative).* As shown in Figure 4, different interviewers describe each job orally using the details of each job presented textually on the screen (to a greater or lesser extent). Again, the words in bold on the transcripts correspond lexically to those on the screen.

In Task 2, cases 1, 2, 3 and 6 describe the details relating to the two jobs presented (bilingual sales manager and the travel agent) as an extended oral turn. However, the interviewer for case 4 presents her partner with two jobs that are not part of the task design: a primary school teacher and secondary school teacher of English and positions herself as a headmistress of the school who is calling by telephone. Similarly, case 5 presents their partner with two jobs: a receptionist night technician (the learner's own words) who does manicures and a personal trainer. This means that cases 4 and 5 have created their own jobs – possibly responding to the textual instruction in the task to "be creative". These cases completely *background* the jobs presented to them textually on the screen by not attending to the job descriptions provided. This result highlights that through mediation learners can decide on the importance of one textual instruction over another, which may be designed to complement each other in the initial task design but which learners can reconfigure with their own agency, depending on the importance or (not) they give to it. Although learners are responding to textual instructions in a series of oral turns they are also responding to the textual prompts (or not) as a conversational resource.

In addition to the text being used as a conversational resource, the learners' navigational trajectories with the audioconferencing tool were found to be different. These trajectories were investigated by analysing the number of seconds before the start of the first turn with their partner (see Appendix E). This was because learners would have needed to read a number of pieces of textual information before they could start the oral task. Reading these textual resources either aloud or silently would have taken time (text about purpose of the task; text about time limit; text warning about efficient use of time and text as lead-in to the sample information). Therefore, if learners were reading (as a response to the screen), evidence of long silences would be apparent in the recordings. Cases 1 and 3 had the longest amount of time before beginning verbal turn-taking. Balaman and Sert (2017) suggest that long silences can point to ongoing orientations to task interface. We also deduce that during these silences, learners are orientating towards or *foregrounding* the screen-based resources and therefore orientating away from or *backgrounding* their peer.

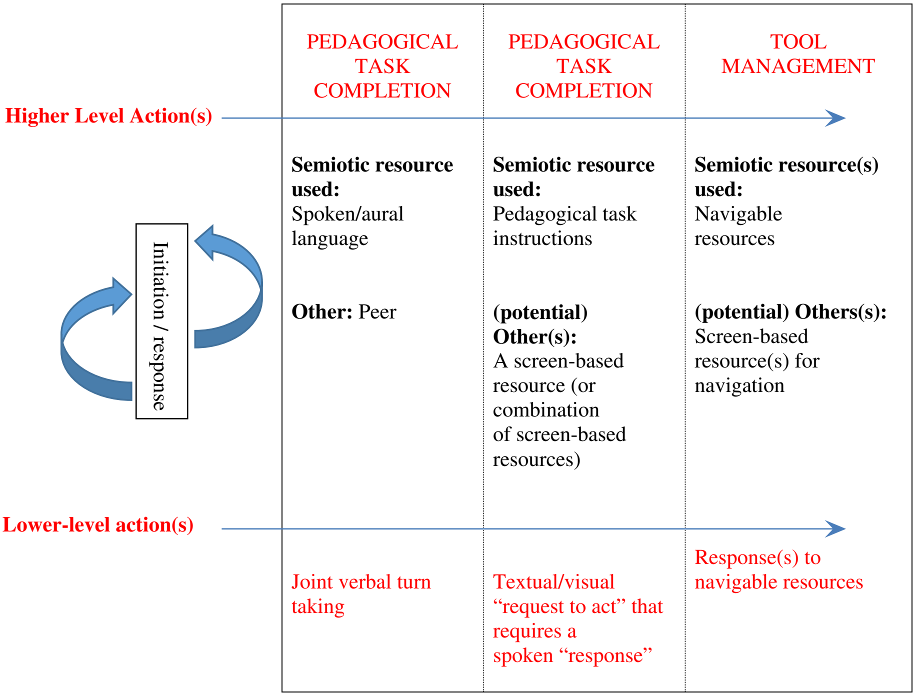
In contrast, in cases 2, 4, 5 and 6, spoken interaction starts almost straight away, indicating that learners had no need to attend to the textual instructions and therefore were *backgrounding* these resources. This may be because they had pre-read them on a previous entry to the tool.

Finally, with respect to the insights pertaining to mediation with navigational resources we understand that learners' non-use of the navigational resources and lack of attention to the pop-up "Time Up!"/"close" suggest that some learners appear to prefer to control the task conditions (controlling their own time and navigational moves). This highlights Norris's (2006, 2016) notion of simultaneous awareness/attention. By not paying attention to certain resources during the process of oral task completion learners are intentionally *backgrounding* their importance as they carry out the task orally and simultaneously.

We now present the following framework that summarises the findings and draws on Norris's (2004, 2006) notion of *higher-level* and *lower-level actions* to highlight how actions in the mediation process can be conceptualised as having two distinct but entwined goals in which the learners' actions appeared to be directed towards Figure 5. The *higher-level actions*, as large-scale actions that learners achieved in the task process, were the completion of the pedagogical task and the management of the Tandem tool. These two higher-level actions were made up of a multiplicity of chained *lower-level actions*. The pedagogical task completion was achieved through the lower-level actions of verbal turns, jointly co-constructed between peers as well as what can be understood as the response (in the form of a series of oral turns) to pedagogical task instructions (e.g. create five questions) that request learners to act.

The higher-level action of (technological) tool management is achieved through learners' use of the navigational resources as a chain of lower-level actions (navigational clicks back and forward), which are both individually (physically) and jointly accomplished and sometimes orally negotiated. The lower-level action of navigation means that learners can reconfigure (navigate

https://doi.org/10.1017/S0958344019000132 Published online by Cambridge University Press



| PEDAGOGICAL TASK COMPLETION | PEDAGOGICAL TASK COMPLETION | TOOL MANAGEMENT |
|---|---|---|
| **Semiotic resource used:** Spoken/aural language | **Semiotic resource used:** Pedagogical task instructions | **Semiotic resource(s) used:** Navigable resources |
| **Other:** Peer | **(potential) Other(s):** A screen-based resource (or combination of screen-based resources) | **(potential) Others(s):** Screen-based resource(s) for navigation |
| Joint verbal turn taking | Textual/visual “request to act” that requires a spoken “response” | Response(s) to navigable resources |

**Figure 5.** The higher-level and lower-level actions (Norris, 2004, 2006) in an audioconferencing task

back and out of the tool and re-enter, not just navigate forward) and repurpose (retrieve task information and instructions without being navigated or pressured by potential pop-ups and timers) the original task and tool design.

The framework highlights that learners are “multitasking” (Knight & Barberà, 2016) as they carry out oral turns and initiate and respond to screen-based resources. Learner agency is also evident in how they attach (or fail to attach) importance to the textual instructions and prompts and in how they repurpose the navigational resources.

Although semiotic mediation for pedagogical task completion appears relatively straightforward (spoken turns as lower-level actions are used to achieve higher-level action of pedagogical task completion), the mediational process is more complex for tool management. It is not only a multimodal process involving different oral turn-taking combinations but also a multisensory process where learners attend to the screen through their sensory (visual) system and motor system as they touch, responding to screen-based resources that also initiate action.

## 5. Conclusion

The aim of the study was to understand how learners carried out mediated action as task completion through and/or with the available semiotic resources in an online spoken interaction task. “Semiotic initiation and response” (Coffin & Donohue, 2014) was found to generally characterise the mediated action for task completion.

The results reveal that audioconferencing as a mode-as-channel of communication is not necessarily “voice but no image” (Yamada, 2009: 820) but that learners can be faced with a variety of visual/textual/navigational screen-based resources as a “multiple ensemble” (Bezemer & Kress, 2016) whereby they use the screen-based resources to carry out the mediated action. Learners were found to mediate their action through the use of task instructions (a textual screen-based "request") and a series of oral turns (as a "response") while also carrying out oral turns with their partner. In addition, learners' encounter with the pedagogical task instructions and navigational resources showed how the sense-making was generally constructed one-sidedly, rather than co-constructed, making the learner solely responsible for the emerging meaning. However, when learners initiate with navigational resources, the computer responded to this by moving them to another page/interface page.

In addition, the avoidance of certain types of experience by some students while speaking (e.g. being navigated, being pressured by time) was noted. For online learners in particular, two of the affordances of online learning are that learners can control their own time and that they have navigational freedom. Therefore, tools that intend to control both of these "affordances" may be a specific problematic area for tool and/or task design because learners' own intentions regarding tool use may be different from those of the designer.

Finally, the terms "videoconferencing", "audioconferencing" and "text-chat" that suggest that CMC is being carried out through monomodal channels, on closer inspection reveal this is increasingly less so (e.g. Lamy, 2006; Liddicoat, 2010). The use of these terms therefore potentially masks over other screen-based modes learners may be faced with such as text in videoconferencing and image in text chat. With respect to future research, more studies are needed that take into account the growing importance of the multiple screen-based resources in language learning. This study confirms that learners are attending to both their peer and the screen. Relying purely on what learners say or do not say is therefore not sufficient to capture how learners may be positioned as semiotic "initiators" and "responders" (Coffin & Donohue, 2014) in relation to the screen.

The study conceptualised screen-based resources as being (potentially) relevant "others" in their support for participants' turn-taking. This led to identifying turns carried through non-verbal means (i.e. visually/textually/through touch) revealing the importance of the screen interface's materiality. The different steps of attending to what learners made relevant, identifying screen-based resources as conversational resources, analysing the screen and carrying out task simulation highlighted the presence of navigational resources as "agents" that could initiate or respond. This helped the researchers attend not only to "the silent visible displays of the hearer work" (Goodwin, 2013: 8) of participants (in oral interaction) but also to the silent visible displays of the screen and the role all of this had in the co-operative social organisation involved in a single shared action. Considering signs and text/discourse as conversational resources (e.g. Kraut et al., 2003) and as part of the mediation with "others" (Raudaskoski, 1999) provided a perspective of a semiotic field or "layer" of action (Goodwin, 2013) that laid the groundwork for showing how resources can also act as or be responded to as conversational agents.

This study has combined different perspectives on the data and as such has provided fuller insight into both human–human and human–computer interaction. However, this is not without limitations. One major limitation is that we cannot see whether learners' orientations are happening in real time, whether they are verbal reports of their previous actions or whether these orientations co-occur with talk, minute by minute. Another limitation is that the research method did not use recordings of the participants' screens, but rather their interactions with the screen were "observed" by (a) listening to the interviewer in combination with (b) the researchers' screenshots that simulated the task process. However, because the aim of the study was to establish whether mediation with screen-based resources occur in the first place, and if so, which ones, rather than to offer a full minute-by-minute account, we take the discourse as a "fingerprint" alongside the triangulation of sources to be robust enough to answer this particular research question. Whereas a combination of eye-tracking technologies with micro-analytic discourse arguably offers a more accurate understanding of the minute-by-minute action, it may be difficult to recruit participants in a context of online learning, and participants may prefer that their behaviours remain undisclosed. In these contexts, approaches such as the one we use may be more fit. The incorporation of learner questionnaires about their behaviours before they record their oral

https://doi.org/10.1017/S0958344019000132 Published online by Cambridge University Press

interaction in the target language (e.g. what they choose to look at on the screen before the task and how they agree on a strategy to present their interaction), as well as think-aloud protocols, would increase validity.

This study offers a number of implications for task design. There is the consideration as to whether screen-based resources *represented* to learners are intended to prompt or shape spoken turns or as a means for the teacher/designer to *communicate* to students (which may draw the learner's attention away from their peer). The "interplay" of modes (Schnotz, 1999) is also important, particularly as one result tentatively suggests that the textual/visual prompts may be able to support an increase in number of turns taken than the number suggested in the textual instructions – a common goal for spoken interaction tasks.

**Ethical statement.** Following the ethical code of our institution (Universitat Oberta de Catalunya) and official regulation of our country (Spain), by this statement we ensure the quality and integrity of our research where all the participants have participated in the research voluntarily with the corresponding informed consent. As authors we have respected the confidentiality and anonymity of the research respondents and we can affirm our research is independent and impartial.

## References

Balaman, U. & Sert, O. (2017) Local contingencies in L2 tasks: A comparison of context-sensitive interactional achievements across two different task types. *Bellaterra Journal of Teaching & Learning Language & Literature*, 10(3): 9–27. https://doi.org/10.5565/rev/jtl3.746

Benson, P. (2015) YouTube as text: Spoken interaction analysis and digital discourse. In Jones, R. H., Chik, A. & Hafner, C. A. (eds.), *Discourse and digital practices: Doing discourse analysis in the digital age*. Oxford: Routledge, 81–96.

Bezemer, J. & Kress, G. (2014) Touch: A resource for making meaning. *Australian Journal of Language and Literacy*, 37(2): 77–85.

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

Canto, S., de Graaff, R. & Jauregi, K. (2014) Collaborative tasks for negotiation of intercultural meaning in virtual worlds and video-web communication. In González-Lloret, M. & Ortega, L. (eds.), *Technology-mediated TBLT: Researching technology and tasks*. Washington: Georgetown University Press, 183–212. https://doi.org/10.1075/tblt.6.07can

Clark, H. H. & Brennan, S. E. (1991) Grounding in communication. In Resnick, L. B., Levine, J. M. & Teasley, S. D. (eds.), *Perspectives on socially shared cognition*. Washington: American Psychological Association, 127–149. https://doi.org/10.1037/10096-006

Coffin, C. & Donohue, J. (2014) *A language as social semiotic-based approach to teaching and learning in higher education* (Language Learning Monograph Series). Hoboken: Wiley-Blackwell. https://doi.org/10.1111/lang.2014.64.issue-s1

Collentine, K. (2009) Learner use of holistic language units in multimodal, task-based synchronous computer-mediated communication. *Language Learning & Technology*, 13(2): 68–87.

Dooly, M. (2018) "I do which the question": Students' innovative use of technology resources in the language classroom. *Language Learning & Technology*, 22(1): 184–217.

Dourish, P., Bellotti, V., Mackay, W. & Ma, C.-Y. (1993) Information and context: Lessons from the study of two shared information systems. In *Proceedings of the Conference on Organizational Computing Systems* (pp. 42–51). Milpitas, CA, 1–4 November. https://doi.org/10.1145/168555.168560

Flewitt, R. (2008) Multimodal literacies. In Marsh, J. & Hallet, E. (eds.), *Desirable literacies: Approaches to language and literacy in the early years* (2nd ed.). London: SAGE, 122–139. https://doi.org/10.4135/9781446279519.n7

Flewitt, R., Hampel, R., Hauck, M. & Lancaster, L. (2013) What are multimodal data and transcription? In Jewitt, C. (ed.), *The Routledge Handbook of Multimodal Analysis* (2nd ed.). London: Routledge, 44–59.

Gilabert, R., Manchón, R. & Vasylets, O. (2016) Mode in theoretical and empirical TBLT research: Advancing research agendas. *Annual Review of Applied Linguistics*, 36: 117–135. https://doi.org/10.1017/S0267190515000112

Goodwin, C. (2013) The co-operative, transformative organization of human action and knowledge. *Journal of Pragmatics*, 46(1): 8–23. https://doi.org/10.1016/j.pragma.2012.09.003

Hampel, R. & Hauck, M. (2006) Computer-mediated language learning: Making meaning in multimodal virtual learning spaces. *The JALT CALL Journal*, 2(2): 3–18.

Hampel, R. (2010) Task design for a virtual learning environment in a distant language course. In Thomas, M. & Reinders, H. (eds.), *Task-based language learning and teaching with technology*. London: Continuum, 131–153.

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

Hauck, M. (2010) The enactment of task design in telecollaboration 2.0. In Thomas, M. & Reinders, H. (eds.), *Task-based language learning and teaching with technology*. London: Continuum, 197–217.

https://doi.org/10.1017/S0958344019000132 Published online by Cambridge University Press

Helm, F. & Dooly, M. (2017) Challenges in transcribing multimodal data: A case study. *Language Learning & Technology*, 21(1): 166–185. https://dx.doi.org/10125/44600

Herring, S. C. (2015) New frontiers in interactive multimodal communication. In Georgakopoulou, A. & Spilioti, T. (eds.), *The Routledge handbook of language and digital communication*. New York: Routledge, 398–402.

Jenks, C. J. (2014) *Social interaction in second language chat rooms*. Edinburgh University Press.

Jewitt, C., Bezemer, J. & O'Halloran, K. (2016) *Introducing multimodality*. London: Routledge.

Knight, J. & Barberà, E. (2016) The negotiation of shared and personal meaning making in spoken interaction tasks. In Pixel (ed.), *Proceedings of the 9th International Conference on ICT for Language Learning* (pp. 248–252). Florence, Italy, 17–18 November. https://conference.pixel-online.net/ICT4LL/acceptedabstracts_scheda.php?id_abs=1994

Kraut, R. E., Fussell, S. R. & Siegel, J. (2003) Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction*, 18(1-2): 13–49. https://doi.org/10.1207/S15327051HCI1812_2

Kress, G. (2003) *Literacy in the new media age*. London: Routledge. https://doi.org/10.4324/9780203299234

Kress, G. & van Leeuwen, T. (2001) *Multimodal discourse: The modes and media of contemporary communication*. London: Edward Arnold.

Lantolf, J. P. (ed.) (2000) *Sociocultural theory and second language learning*. Oxford: Oxford University Press.

Lamy, M.-N. (2006) Multimodality in second language conversations online: Looking for a methodology. In *Proceedings of the Third International Conference on Multimodality*. Pavia, Italy, 385–403.

Liddicoat, A. J. (2010) Enacting participation: Hybrid modalities in on-line video conversation. In Develotte, C., Kern, R. & Lamy, M.-N. (eds.), *Décrire la conversation en ligne: Le face à face distanciel*. Lyon: ENS Éditions, 37–50. http://catalogue-editions.ens-lyon.fr/fr/livre/?GCOI=29021100952500

Liou, H.-C. (2011) Blogging, collaborative writing, and multimodal literacy in an EFL context. *WorldCALL: International perspectives on computer-assisted language learning*. New York: Routledge, 3–18.

Mondada, L. (2007) Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers. *Discourse Studies*, 9(2): 194–225. https://doi.org/10.1177/1461445607075346

Mondada, L. (2013) Embodied and spatial resources for turn-taking in institutional multi-party interactions: Participatory democracy debates. *Journal of Pragmatics*, 46(1): 39–68.

Norris, S. (2004) *Analyzing multimodal interaction: A methodological framework*. London: Routledge. https://doi.org/10.4324/9780203379493

Norris, S. (2006) Multiparty interaction: A multimodal perspective on relevance. *Discourse Studies*, 8(3): 401–421. https://doi.org/10.1177/1461445606061878

Norris, S. (2016) Concepts in multimodal discourse analysis with examples from video conferencing. *Yearbook of the Poznan Linguistic Meeting*, 2(1): 141–165. https://doi.org/10.1515/yplm-2016-0007

Oben, B. & Brône, G. (2015) What you see is what you do: On the relationship between gaze and gesture in multimodal alignment. *Language and Cognition*, 7(4): 546–562. https://doi.org/10.1017/langcog.2015.22

Raudaskoski, P. (1999) *The use of communicative resources in language technology environments: A conversation analytic approach to semiosis at computer media*. University of Oulu, unpublished PhD. http://vbn.aau.dk/ws/files/72280035/pirkkosphd.pdf

Rossolatos, G. (2015) Taking the "multimodal turn" in interpreting consumption experiences. *Consumption Markets & Culture*, 18(5): 427–446. https://doi.org/10.1080/10253866.2015.1056167

Sacks, H., Schegloff, E. A. & Jefferson, G. (1974) A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4): 696–735.

Schnotz, W. (1999) Introduction. *European Journal of Psychology of Education*, 14(2): 163–165. https://doi.org/10.1007/BF03172963

Tudini, V. (2014) Conversation analysis of computer-mediated interactions. In Chapelle, C. (ed.), *The encyclopedia of applied linguistics*. Hoboken: John Wiley & Sons, 1–7. https://doi.org/10.1002/9781405198431.wbeal1456

van Lier, L. (2000) From input to affordance: Social-interactive learning from an ecological perspective. In Lantolf, J. (ed.), *Sociocultural theory and second language learning*. Oxford: Oxford University Press, 245–259.

Vetter, A. & Chanier, T. (2006) Supporting oral production for professional purposes in synchronous communication with heterogenous learners. *ReCALL*, 18(1): 5–23. https://doi.org/10.1017/S0958344006000218

Vygotsky, L. S. (1981) The instrumental method in psychology. In Wertsch, J. V. (ed.), *The concept of activity in Soviet psychology*. Armonk: Sharpe, 134–143.

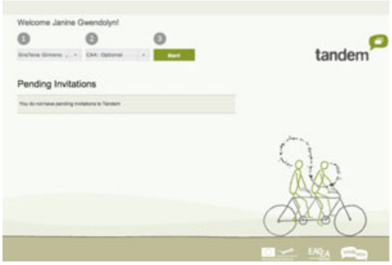
Wertsch, J. V. (1994) The primacy of mediated action in sociocultural studies. *Mind, Culture, and Activity*, 1(4): 202–208.

Wertsch, J. V. (1998) *Mind as action*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195117530.001.0001

Yamada, M. (2009) The role of social presence in learner-centered communicative language learning using synchronous computer-mediated communication: Experimental study. *Computers & Education*, 52(4): 820–833. https://doi.org/10.1016/j.compedu.2008.12.007

## Appendix A. Example screenshots of reconstructed-task simulation by researcher

1.

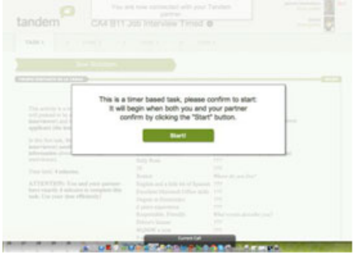


Welcome Janine Gwendolyn!

Pending Invitations

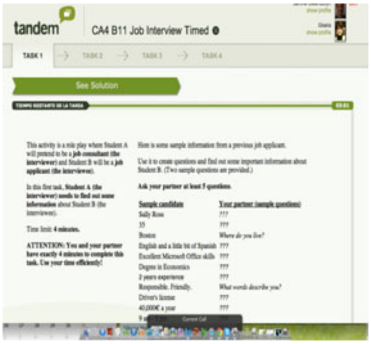
You do not have pending invitations to Tandem

2.



This is a timer based task, please confirm to start. It will begin when both you and your partner confirm by clicking the "Start" button.

3.



This activity is a role play where Student A will pretend to be a **job consultant (the interviewer)** and Student B will be a **job applicant (the interviewee).**

In this first task, **Student A (the interviewer) needs to find out some information** about Student B (the interviewee).

Time limit: **4 minutes.**

**ATTENTION: You and your partner have exactly 4 minutes to complete this task. Use your time efficiently!**

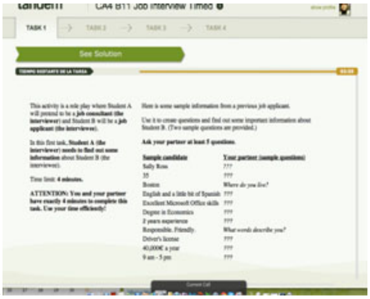
Here is some sample information from a previous job applicant.

Use it to create questions and find out some important information about Student B. (Two sample questions are provided.)

**Ask your partner at least 5 questions.**

| Sample candidate | Your partner (sample questions) |
|---|---|
| Sally Ross | ??? |
| 35 | ??? |
| Boston | *Where do you live?* |
| English and a little bit of Spanish | ??? |
| Excellent Microsoft Office skills | ??? |
| Degree in Economics | ??? |
| 2 years experience | ??? |
| Responsible. Friendly. | *What words describe you?* |
| Driver's license | ??? |
| 40,000€ a year | ??? |
| 9 | |

4.



This activity is a role play where Student A will pretend to be a **job consultant (the interviewer)** and Student B will be a **job applicant (the interviewee).**

In this first task, **Student A (the interviewer) needs to find out some information** about Student B (the interviewee).

Time limit: **4 minutes.**

**ATTENTION: You and your partner have exactly 4 minutes to complete this task. Use your time efficiently!**

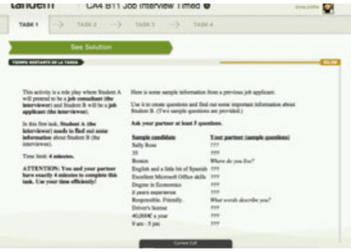
Here is some sample information from a previous job applicant.

Use it to create questions and find out some important information about Student B. (Two sample questions are provided.)

**Ask your partner at least 5 questions.**

| Sample candidate | Your partner (sample questions) |
|---|---|
| Sally Ross | ??? |
| 35 | ??? |
| Boston | *Where do you live?* |
| English and a little bit of Spanish | ??? |
| Excellent Microsoft Office skills | ??? |
| Degree in Economics | ??? |
| 2 years experience | ??? |
| Responsible. Friendly. | *What words describe you?* |
| Driver's license | ??? |
| 40,000€ a year | ??? |
| 9 am - 5 pm | ??? |

5.



This activity is a role play where Student A will pretend to be a **job consultant (the interviewer)** and Student B will be a **job applicant (the interviewee).**

In this first task, **Student A (the interviewer) needs to find out some information** about Student B (the interviewee).

Time limit: **4 minutes.**

**ATTENTION: You and your partner have exactly 4 minutes to complete this task. Use your time efficiently!**

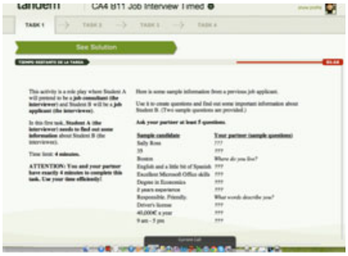
Here is some sample information from a previous job applicant.

Use it to create questions and find out some important information about Student B. (Two sample questions are provided.)

**Ask your partner at least 5 questions.**

| Sample candidate | Your partner (sample questions) |
|---|---|
| Sally Ross | ??? |
| 35 | ??? |
| Boston | *Where do you live?* |
| English and a little bit of Spanish | ??? |
| Excellent Microsoft Office skills | ??? |
| Degree in Economics | ??? |
| 2 years experience | ??? |
| Responsible. Friendly. | *What words describe you?* |
| Driver's license | ??? |
| 40,000€ a year | ??? |
| 9 am - 5 pm | ??? |

6.



This activity is a role play where Student A will pretend to be a **job consultant (the interviewer)** and Student B will be a **job applicant (the interviewee).**

In this first task, **Student A (the interviewer) needs to find out some information** about Student B (the interviewee).

Time limit: **4 minutes.**

**ATTENTION: You and your partner have exactly 4 minutes to complete this task. Use your time efficiently!**

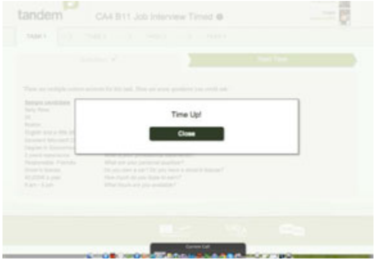
Here is some sample information from a previous job applicant.

Use it to create questions and find out some important information about Student B. (Two sample questions are provided.)

**Ask your partner at least 5 questions.**

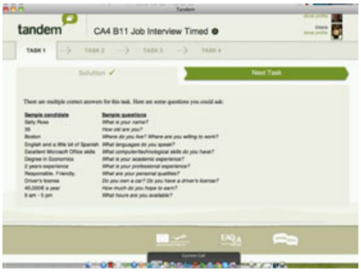
| Sample candidate | Your partner (sample questions) |
|---|---|
| Sally Ross | ??? |
| 35 | ??? |
| Boston | *Where do you live?* |
| English and a little bit of Spanish | ??? |
| Excellent Microsoft Office skills | ??? |
| Degree in Economics | ??? |
| 2 years experience | ??? |
| Responsible. Friendly. | *What words describe you?* |
| Driver's license | ??? |
| 40,000€ a year | ??? |
| 9 am - 5 pm | ??? |

7.



Time Up!

8.



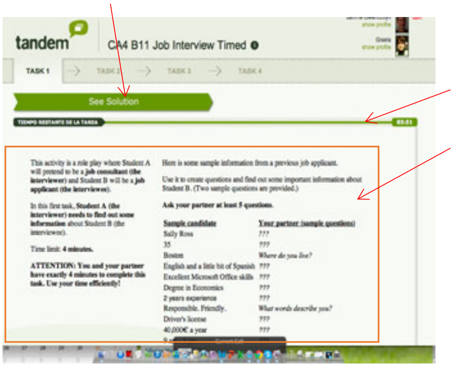
There are multiple correct answers for this task. Here are some questions you could ask:

| Sample candidate | Sample questions |
|---|---|
| Sally Ross | *What is your name?* |
| 35 | *How old are you?* |
| Boston | *Where do you live? Where are you willing to work?* |
| English and a little bit of Spanish | *What languages do you speak?* |
| Excellent Microsoft Office skills | *What computer/technological skills do you have?* |
| Degree in Economics | *What is your academic experience?* |
| 2 years experience | *What is your professional experience?* |
| Responsible. Friendly. | *What are your personal qualities?* |
| Driver's license | *Do you own a car? Do you have a driver's license?* |
| 40,000€ a year | *How much do you hope to earn?* |
| 9 am - 5 pm | *What hours are you available?* |

https://doi.org/10.1017/S0958344019000132 Published online by Cambridge University Press

## Appendix B. Labelling of the multimodal ensemble of interface pages (Task 1)

### Visual/textual/navigable resources



This activity is a role play where Student A will pretend to be a **job consultant (the interviewer)** and Student B will be a **job applicant (the interviewee)**.

In this first task, **Student A (the interviewer) needs to find out some information** about Student B (the interviewee).

Time limit: **4 minutes.**

**ATTENTION: You and your partner have exactly 4 minutes to complete this task. Use your time efficiently!**

Here is some sample information from a previous job applicant.

Use it to create questions and find out some important information about Student B. (Two sample questions are provided.)

**Ask your partner at least 5 questions.**

| Sample candidate | Your partner (sample questions) |
|---|---|
| Sally Ross | ??? |
| 35 | ??? |
| Boston | *Where do you live?* |
| English and a little bit of Spanish | ??? |
| Excellent Microsoft Office skills | ??? |
| Degree in Economics | ??? |
| 2 years experience | ??? |
| Responsible. Friendly. | *What words describe you?* |
| Driver's license | ??? |
| 40,000€ a year | ??? |

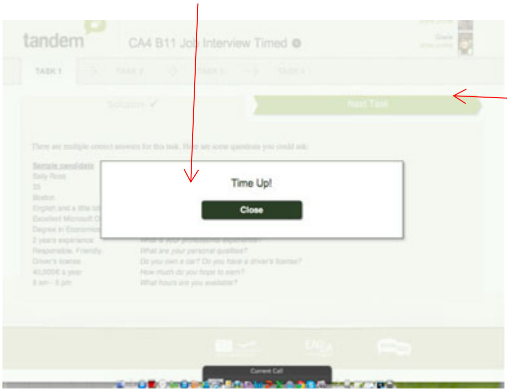
**Visual/textual resources:**
Information about task sequence

**Visual/textual/navigational**
"See Solution"

**Visual resources (dynamic)**
Timer, pop-up

**Textual/visual resources:**
Information about roles
Information about the communicative purpose
Instructions
Prompts
Information: time limit with warning

Time Up! Pop-up (visual/textual)/close (navigational)



There are multiple correct answers for this task. Here are some questions you could ask.

| Sample candidate | |
|---|---|
| Sally Ross | |
| 35 | |
| Boston | |
| English and a little bit | |
| Excellent Microsoft O | |
| Degree in Economics | |
| 2 years experience | *What is your professional experience?* |
| Responsible. Friendly. | *What are your personal qualities?* |
| Driver's license | *Do you own a car? Do you have a driver's license?* |
| 40,000€ a year | *How much do you hope to earn?* |
| 9 am – 5 pm | *What hours are you available?* |

**Visual/textual/navigational**
"Next task" (phased out) and "start" button

## Appendix C. Analysis of screen-based resources

| | |
|---|---|
| No screenshot | Waiting for confirmation |
| **Textual/visual resources (static):** Screenshots 3–6 | INTERFACE PAGE 1 ON SCREEN |
| | – "Task 1" foregrounded (Task, 2, 3 and 4 backgrounded) |
| | – about roles or Student A or B |
| | – about purpose of the task |
| | – about time limit |
| | – warning about efficient use of time |
| | – lead-in to sample information |
| | – create own questions + symbol? |
| | – instruction to ask the minimum of 5 questions |
| | – Text from sample candidate |
| | – Textual sample questions; e.g. "Where do you live?" |

https://doi.org/10.1017/S0958344019000132 Published online by Cambridge University Press

| | |
|---|---|
| **Visual (dynamic):**<br>Screenshots 3–6 | – The visual timer counting the seconds left and changes colour with "*Tiempo Restante de la Tarea*" (Time left for task) banner |
| **Visual/textual/navigable:**<br>Screenshots 3–6 | – "See Solution" banner |
| **Visual/textual/navigable (dynamic)**<br>Screenshots 7 | SCREEN (INTERFACE PAGE 1 BACKGROUNDED)<br>– A pop-up sign indicating "Time up!"and "Close" |
| Screenshot 8 | INTERFACE PAGE 2 "SOLUTIONS PAGE" |
| **Textual (static)**<br>Screenshot 8 | – information that there are multiple correct answers |
| Screenshot 8 | – information of sample candidate and sample questions |
| **Visual/textual/Navigable**<br>Screenshot 8 | – "Next task" banner |
| Screenshot 9 | SCREEN (INTERFACE PAGE 2 BACKGROUNDED) |
| **Visual/textual/navigable**<br>Screenshot 9 | – information that it is a timed task<br>request to start by clicking the "Start" button |
| Screenshot 10–12 | INTERFACE PAGE 3 |
| | – "Task 2" foregrounded<br>(Task 1, 3 and 4 backgrounded) |
| **Textual/visual**<br>Screenshot 10–12 | – information that two jobs are available and the need to explain them |
| | – information that partner should ask 3 questions and say which they prefer |
| | – information about time limit |
| | – instruction to describe the two jobs to partner, answer partner's questions and be creative if you don't know the answers |
| | Text describing two jobs |
| **Visual (dynamic):**<br>Screenshot 10–12 | – The visual timer counting the seconds left and changes colour with "Tiempo Restante de la Tarea" banner |
| **Visual/textual/navigable (dynamic)**<br>Screenshot 13 | SCREEN (INTERFACE PAGE 2 BACKGROUNDED)<br>– A pop-up sign indicating "Time up!"and "Close" |
| **Visual/textual/navigable** | – "Next Task" banner |

## Appendix D. Time on Tasks 1 and 2

| Case | Time required to complete Tasks 1 and 2 | Time taken by cases |
|---|---|---|
| Case 1 | 11 minutes for all cases | 14:33 |
| Case 2 | | 8:07 |
| Case 3 | | 11:48 |
| Case 4 | | 3:33 |
| Case 5 | | 6:20 |
| Case 6 | | 9:31 |

https://doi.org/10.1017/S0958344019000132 Published online by Cambridge University Press

## Appendix E. Number of seconds before start of talk

| | |
|---|---|
| Case 1 | 23 seconds |
| Case 2 | 1 second |
| Case 3 | 28 seconds |
| Case 4 | 1 second |
| Case 5 | 1 second |
| Case 6 | 3 seconds |

### About the authors

**Janine Knight**, holds a doctorate in education and e-learning and is currently a teacher of English and teacher trainer at the Universitat Internacional de Catalunya, Spain. Her research activity is focused on learner agency in computer-assisted language learning (CALL) and task-based learning scenarios to develop spoken interaction.

**Melinda Dooly** holds a Serra Húnter fellowship as senior lecturer at the Universitat Autònoma de Barcelona. Her principal research addresses technology-enhanced project-based language learning, intercultural communication and 21st century competences in teacher education. She is lead researcher of GREIP: Grup de Recerca en Ensenyament i Interacció Plurilingües (Research Centre for Teaching and Plurilingual Interaction).

**Elena Barberà** holds a doctorate in psychology from the University of Barcelona (1995). A full professor, she is currently head of the doctoral program in ICT and education of the Open University of Catalonia in Barcelona (Spain). Her research activity is specialised in the area of educational psychology, relating in particular to knowledge-construction processes and educational interaction in e-learning environments, evaluating educational quality and assessing learning, distance learning using ICT and teaching and learning strategies.

Author ORCiD. Janine Knight, https://orcid.org/0000-0001-7438-3698
Author ORCiD. Melinda Dooly, https://orcid.org/0000-0002-1478-4892
Author ORCiD. Elena Barberà, https://orcid.org/0000-0002-9315-8231

https://doi.org/10.1017/S0958344019000132 Published online by Cambridge University Press