But the computer say me the time is up: the shaping of oral turns mediated *with* and *through* the screen

Janine Knight[a,b], Melinda Dooly[c] and Elena Barberà[d]

[a]Institute of Multilingualism, Universitat Internacional de Catalunya, Barcelona, Spain; [b]Universitat Oberta de Catalunya University, Barcelona, Spain; [c]Department of Language and Literature Education and Social Science Education, Universitat Autònoma de Barcelona, Bellaterra, Spain; [d]Doctorate School, Universitat Oberta de Catalunya University, Barcelona, Spain

ABSTRACT

Research on mode in computer-mediated communication and language learning has primarily focused on mode-as-channel of communication such as audio- or videoconferencing. However, increasingly sophisticated technological tools now facilitate communication in multiple ways so that learners can convey and respond to peers and screen-based resources, both orally, visually and/or through touch with the screen. By highlighting learner-to-learner turns as well as screen related turns, this paper looks at how learners' oral meaning making is shaped through both verbal and non-verbal resources. This can provide a better understanding of how the interplay between modes and resources on interfaces might be harnessed to increase learners' oral turns and identify potential interface-related difficulties learners might face. Audio recordings of six dyads are analysed using discourse analysis, with notions from conversational analysis, alongside interface screenshots. Results reveal that screen-based resources become: 1) embedded or modified in oral turns; 2) resources to initiate and support oral turns; 3) diverse topics of talk. It was also found that learners orientate towards some resources as 4) agentive turn-takers (e.g. pop-ups, to initiate an action). Such multimodal experience reveal how peer-to-peer talk can occasionally resemble a multi-party encounter whereby some resources can act as participants in the interaction. It is proposed that a fuller understanding of this interplay can help teachers and designers optimize computer-mediating communicative language learning tasks.

KEYWORDS

Computer-mediated communication; multimodal turn-taking; screen-based resources; semiotic resources; spoken interaction; task based language learning

1. Introduction

Screen-based visual elements play an increasingly prominent role in modern society (Liou, 2011) such as interactive banking, getting medical

appointments etc.; all of which can now be considered the norm. This salient role is also becoming evident in online language education. Online language learners carrying out spoken interaction tasks can use audioconferencing or videoconferencing tools for computer-mediated communication (henceforth CMC). During these interactions learners face the possibility of a 'multimodal ensemble' (Bezemer & Kress, 2016) on the screen. In any session, learners may be attending to the semiotic resources provided by their peers (e.g. aural inputs, gesture and gaze) and/or semiotic resources from the screen (screen-based resources) (e.g. text, visuals, audio). These may also include screen-based navigational resources that learners need to engage with physically through clicking or tapping (in tablets, for instance). In short, the 'interplay of modes' (Schnotz, 1999) of potential resources in an online language learning environment is highly complex and requires considerable effort on behalf of the learner.

In order to understand the complexity of all the modes involved in meaning creation (Calvo-Ferrer, Melchor-Couto, & Jauregi, 2016), such as the scenario described, both learner processes and outcomes of spoken interaction tasks seem increasingly in need of a multimodal lens: a lens that not only considers verbal and non-verbal (textual, visual, tactile/ *kin-aesthetic*) communication but also embodied and non-embodied communication (emanating from humans *and* from computers). However, there a lack of research on the impact of multimodal communication in online language classrooms (Hampel & Stickler, 2012) as well as the role of different mode in language learning(Gilabert, Manchón, & Vasylets, 2016). Furthermore, within the field of second language acquisition (henceforth SLA), the predominant way to analyse such a scenario has been by focusing on learners' meaning making through purely verbal means, highlighting the 'lingual bias' (Block, 2013), whereas, as aforementioned, daily communication is increasingly more complex and requires multimodal communicative competences. It follows, therefore, that a shift in how meaning making is both conceptualised and analysed might take into account the screen that is present and the learners' potential to initiate and/or respond, both orally and physically, to screen-based resources while interacting with their peers, thus moving away from a language-only focus. Most studies in SLA have predominantly focused on mediation using spoken language as a semiotic resource *through* the screen by focusing on different channels of communication (or bi and tri-channels) and its/their impact on human-human spoken interaction. This study focuses on mediation with spoken language as a semiotic resource *through* the screen as well as the external semiotic resources *on* the screen. For language teachers, this is imperative to take into

consideration. As the CEFRL Companion explains: language learners aiming for the highest proficiency (C levels) of online interaction, should be able to 'adapt his/her register and interaction style according to the virtual space he/she is in, adjusting his/her language appropriately to make his communication more effective'. This includes accommodating to 'multi-modal phenomenon' (Council of Europe (COE),), 2018, p. 96).

2. Theoretical framework

2.1. Computer mediated communication for spoken interaction

CMC tools offer practice for speaking and human spoken feedback opportunities either with other L2 learners or L1 speakers (Blake, 2016). Synchronous CMC tools can facilitate videoconferencing e.g. Adobe Connect, Big Blue Button, Blackboard Collaborate, Skype, Google Hangout, Zoom and typically allow learners to exchange video, images, and text in real time and are important tools for developing L2 speaking practice (Blake, 2016). In addition, audioconferencing and audiographic conferencing tools may offer the same affordances but without the added presence of the teacher's or classmate's face and body. Some tools e.g. *netmeeting*, *Tandem* from the SpeakApps project and computer systems e.g. *Lyceum,* allow for audioconferencing to be a channel with 'voice, but no image' (Yamada, 2009). However, learners can indeed face screen-based (non-embodied) image and/or other content with audioconferencing tools. This 'input' may include content from an interface(s) pertaining to (inter)net-based or non net-based sources.

2.1.1. Screen-based resources and turn-taking in audio and video CMC for spoken interaction

Levy and Kennedy (2004) study highlighted the possibilities for screen sharing using the audioconferencing tool *netmeeting*. Learners could share the same screen while browsing the net and talking simultaneously. Non networked-based audioconferencing tools, such as the *Tandem* tool (which this current study focuses on), facilitated the distribution of different (or same) screens to learners so that screen-based textual instructions and navigational resources may be the same or different (Knight & Barberà, 2016). In videoconferencing, screens can be minimized, meaning that the interlocutor's face can 'disappear'. Learners hear sound from the video (as in audioconferencing) but can in fact see net-based content (internet pages). An example of this is outlined in Guth and Helm (2012) study that used Skype audio and video for telecollaboration.

Oral turns and turn-taking has been a focus in many studies on CMC for spoken interaction, focusing on speech as the mode used *through* the screen (see Jenks, 2014 for an overview of methodologies, studies and authors). However, because this current study is focused on learners' mediation, (including possible turn-taking) *with* the screen through the use of various modes while simultaneously carrying out oral interaction with their peers *through* the screen. Or as the CEFRL Companion (COE, 2018) describes, learners are expected to habituate to online register and interaction appropriate to the virtual space, including appropriate bridging, or mediation, of all the resources needed for effective communication. Therefore, we draw on studies from the field of SLA and CMC (audio and video) that have drawn attention to both of these two foci for mediation, namely studies by Lamy, 2006; Liddicoat, 2010; Knight & Barberà, 2016 and Knight, Dooly, & Barberà, 2020. We also take into consideration some pertinent studies from the literature on Human-Computer Interaction (henceforth HCI). This is in order to gain insight into how oral turns may be shaped by screen-based resources while participants interact orally, as well as to gain insight into the non-verbal turns that may be occurring with screen-based resources.

In SLA, some refer to audioconferencing as 'voice, but no image' (Yamada, 2009). However, a number of the above studies on CMC highlight the screen-based images and texts as well as other resources (or modes) that do not necessarily pertain to another person, such as gaze and gesture, while speaking. For instance Lamy's (2006) study, which used the notion of sequentiality from Conversation Analysis, focused turn-taking not only in written and spoken interaction, but also turn-taking mediated through and with visual resources such as icons, images, colours and shapes. She found that when more than one modality was available, turns were managed in a hybrid mode across speech and text so that turns were constructed multimodally (spoken and written turns), therefore operating transmodally (across modes). Similarly, Knight and Barberà's study (2016) on joint and negotiated meaning construction using a non net-based audioconferencing tool (*Tandem*) showed that language learners were multi-tasking in spoken language tasks using different screen-based resources (spoken language, a photo, navigational resources) to achieve different smaller purposes within a task. In addition, Knight, Dooly and Barberà's study (2020) found that learners took oral turns with peers *and* with the various screen-based resources (textual instructions, pop-ups, navigational resources) during a spoken interaction task, highlighting how turn-taking could be shared multimodally with screen-based resources. Furthermore, one result suggested that the amount of textual/visual prompts may be able to support an increase in

number of oral turns taken by learners over the number of turns requested in the textual instructions thus highlighting the importance of the 'interplay between modes'(Schnotz, 1999).

Turns made through human-to-human interaction as well as turns made at specific points through human-computer interaction was also noted by Liddicoat (2010) in his study on videoconferencing, highlighting the 'non-embodied external resources' used by learners in turn-construction. For instance, turn initiations must first secure the attendance of a non-present co-participant, which may be achieved through technology, namely a message via the computer (screen): 'Andrew wants to have a video conversation' which is neither spoken or written by Andrew but is initiated by him with technology. The response by his partner is either a choice to press 'respond' or 'refuse'. This sequence resembles verbal summons-answers sequences such as 'hey' or naming; or technological equivalents of a telephone ringing. Online conversation involves a multimodal achievement of openings, interruptions and closings which, following Tudini (2014, p. 3), 'needed to be both technologically and interactionally accomplished due to constraints of the medium'. In summary, micro-analytical approaches have shown that interactional competence are not limited to verbal or gestural, in online or digital (cellphones, etc.), communicative competence includes the mediation of multiple resources to accomplish online 'collaborative or transactional exchanges' (COE, 2018, p. 99).

Kraut, Fussell, and Siegel (2003) considered the ways in which visual information was used as a conversational resource in the accomplishment of collaborative physical, (non-language learning) tasks and the effects of this visual information on performance and on conversational strategies. They noted that although most systems (e.g. videoconferencing, electronic mail, audioconferencing) are designed to support group activities that can be performed without reference to the external spatial environment (e.g. decision making) they highlight that observational studies of physical collaboration suggest that people's speech and actions are intricately related. Their comparative study of collaborative bike repairing using video and audioconferencing systems and the utterances that occur in the process, relate to different message types. *Procedural* (instructions furthering task completion), *Task status* (state of task or objects in the task), *Referential* (pertaining to the identifications or location of task objects) and *Internal state* (Intentions, knowledge, emotions etc.).

Up until now our literature review for CMC for spoken interaction has focused predominantly on CMC and task-based SLA for spoken interaction, non-language learning CMC studies in which screen-based

resources appear to form turns and pertinent studies from the HCI literature. We now expand our review further to focus on non-verbal, turn-taking with screen-based resources and studies that have considered screen-based resources beyond the lingual. This is in order to form a framework for analysing verbal and non-verbal discourse in relation to turns and turn-taking.

2.2. A Framework for analysing verbal and non-verbal turns and turn-taking

The analysis of discourse in CMC has often been carried out using 'traditional' tools such as discourse analysis or conversational analysis (Benson, 2015). Although historically, spoken interaction has been a central focus through the work of two major schools, namely conversational analysis (CA) (Sacks, Schegloff, & Jefferson, 1974) and the Birmingham school of discourse analysis (Sinclair & Coulthard, 1975), some have queried how far sets of principles from these schools could be applied to spoken interaction through CMC. However, CMC researchers have largely relied on these principles often adapting them (Benson, 2015).

In this study, we build on Benson's (2015) approach of using discourse categories of 'exchange', 'turn', 'move' and 'act', following Coulthard (1985) and Stenström (1994). In Benson's (2015) study, a turn is a basic unit of interactional analysis, and turns are framed by the affordances of the interface page, which governs how users' contributions will appear on the page. If a user contributed semantic content it counted as a turn (e.g. writing a comment and uploading it or clicking a 'like' action). According to Benson (2015) 'in order for a turn to be interactive, this action, whatever its form, must somehow be linked to an action performed by another user' (2015, p. 88).

Furthermore, because turn-taking with a minimal initiation/response is one way that 'an encounter' (Raudaskoski, 1999) may occur, leading to semiotic mediation or semiosis, the main focus of analysis is on the interpretative work of learners in their 'encounters with others' (Raudaskoski, 1999). Raudaskoski (1999) identifies how the 'other' has been generally approached as 1) text or discourse, 2) a social entity and 3) a sign. These three conceptualisations are used to understand the data. Raudaskoski's (1999) study is relevant to this current study because she studied interactions with computers whilst humans talk and identified how humans orientated to them during the talk, whether as 'signs' or as 'entities' for example.

Apart from 'turn', another conceptualisation used in the analysis is the notion of the 'other'. By focusing on language as an action 'by somebody

to do various things in the world (language understood as a present event, with focus on the effects and interaction)' (Raudaskoski, 1999, p. 40), the notion of 'other' becomes a key element in discourse analysis (see speech act theory (Austin, 1962; Searle, 1969 and Grice, 1975). However, 'other' has been conceptualized quite differently, according to different fields and applications of the term. One general approach sees the 'other' as a social entity. This usually implies agency where the whole of society works through its agents' interactions with each other (Raudaskoski, 1999). Therefore, any encounter can be regarded as two or more social agents meeting. Raudaskoski (1999) offers an example of this whereby the computer is referred to as 'this one' as in 'this one speaks American too', as another social entity. However, according to Raudaskoski (1999), the spatio-temporal meaning construction is one-sided, i.e. the human user is responsible for constructing the meaning on the basis of the designer's represented turns. Another approach conceptualises 'the other' as sign and in the field of semiotics, the relation between signs and meaning has been the main focus.

This focus has been picked up an underlying principle to support the analysis of turns and turn-taking with respect to potential screen-based 'others' in the framework. This application echoes the premise that 'modes do not exist without social actors utilising them in some way' (Jewitt, Bezemer, & O'Halloran, 2016, p. 115) to orchestrate the interaction and for 'orientation' purposes. Benson (2015) employed the notion of 'orientation' from Conversational Analysis to understand moves that he conceptualised as 'responses' on a YouTube page (e.g. a video response, 'like/'dislike' icons or written 'comments') to an 'initiation' (uploading a video). From discourse, Sacks et al. (1974) a founder of CA, proposes that by looking at subsequent turns a researcher can ground an analysis of previous ones.

Oral turn-taking as a form of knowledge construction (Sacks, Schegloff, & Jefferson, 1974) has been focused on in SCMC (González-Lloret, 2011; Jenks, 2014) as an analysis of turn sequences can reflect how knowledge is constructed between people.Turn-taking is relevant for SLA because both socio-cultural perspectives (Lantolf, 2000) and cognitive perspectives (de la Colina & García Mayo, 2007; Gass & Mackey, 2006) on how languages are learnt posit that oral interaction is the key for learning languages. While cognitive approaches have focused on quantity of turns through the concept of 'Negotiation for Meaning', whereby more oral turns equates to more interaction taking place, learners' purposes for turn-taking has also been studied with a focus on speech acts and discourse funcions such as jokes and off task-discussions (e.g. Collentine, 2009).

Taking this into consideration, a framework for analysing verbal and non-verbal turns and turn-taking from a multimodal perspective, that

builds on the work of semioticians and conversation analysts, might contribute to L2 learning in a number of ways. It would be useful in order to give teachers and designers insight as to whether the 'interplay between modes' (Schnotz, 1999) and resources can facilitate more or less oral interaction, as suggested in Knight et al. (2020) study. For example, interplays involving different use of colours, space or to identify if the presence of visual prompts such as examplesinstead of textual promptsin the form ofinstructions are more effective in scaffolding more turns- a primary concern in cognitivist approaches to SLA. It would also help teachers to understand learners' talk more holistically, that may emerge beyond the talk expected in relation to the pedagogical task- a focus within sociocultural perspectives to SLA. In addition, such a focus might highlight any difficulties that students face when 'multitasking' as they attend to both the screen-based resources and their partner while they converse – a concern that conceivably would help online task designers.

3. Purpose of study and research questions

The purpose of this study is to understand how oral turns are shaped by learners' encounters with screen-based resources and with what by focusing on how this is made relevant in their oral turns. In addition, the study aims at understanding how learners may initiate or respond through non-verbal semiotic and screen-based resources while they interact orally with their peer.

RQ 1

How are learners' oral turns shaped by encounters with the available screen-based resources in an online task?

RQ 2

How do learners initiate or respond to other screen-based resources and if so, what does this tell us about turn-taking where screen-based resources are considered as potential 'others'?

4. Participants and context

The participants were students in an English as a Foreign Language class, part of their degree programme at a 100% virtual university. The learners were in a B2.1 level on the CEFR (upper intermediate). There were 12 adult students: 2 male and 10 female, 26-55 years old, engaged in a virtual synchronous peer-to-peer speaking role-play task. These synchronous tasks occurred twice in a course, over four months, and while some students may have been familiar with the technology from study in

previous years, others may not have been.Students were bilingual (Catalan and Spanish) with English as an additional language. The students form 6 dyads as case studies. Data sources include approximately 97 minutes of peer-to-peer recorded oral conversations. In addition, screenshots of the task were taken randomly by the main researcher Samples of these can be seen in Appendix A.

The task was a role-play activity in which learners had to take turns being the interviewer and the interviewee. Only audio was available, so they could not see each other. The role-play task was sub-divided into two tasks (Task 1 and Task 2) which was then repeated as learners changed roles thus becoming Task 4 and 5. Task 1 required the learners to ask questions in the role of interviewer and the peer to respond as the interviewee, as can been seen in the screenshot, Figure 1. Interviewers were required to ask minimally 5 questions and prompted to elicit some information from the interviewees. Sample textual screen-based information was provided. In Task 2, interviewers were instructed to describe two jobs to their partner. Both students had a description of the task but only the interviewer, assigned with the role of student B by the Tandem tool, saw the questions and information regarding the sample candidate. This can be seen in Figure 1. At the same time, student A, assigned with the role of interviewee, saw a different interface. This interface presented them with textual information about the roles, the time limit, instructions to answer their partner's questions and to answer in full sentences using the language for a job. This can be seen in Figure 2.

The tasks were timed with a timer showing how much time learners had left and a pop-up icon emerged when the time was up.

After the two tasks were completed (indicated by the banner 'TASK 1′ and 'TASK 2′ in Figure 1 the tasks were repeated and the peers changed rolesindicated by the banners 'TASK 3′ and 'TASK 4′ in Figures 1 and 2. Learners had to navigate from one task to the next.

One of the features of the Tandem tool was that the tool could assign different screen content to different students. This meant that each student saw a different screen once the roles were allocated to them. The textual instructions related to the pedagogical task differed between students as did the the textual scaffolds to support them in carrying out their roles successfully. Therefore, the personal and job details, for example, were only seen by the student playing the role of the interviewer.

5. Instruments and analysis procedure

For the analytical procedure, the main researcher compared the transcripts with the screenshots to check whether there was coherence

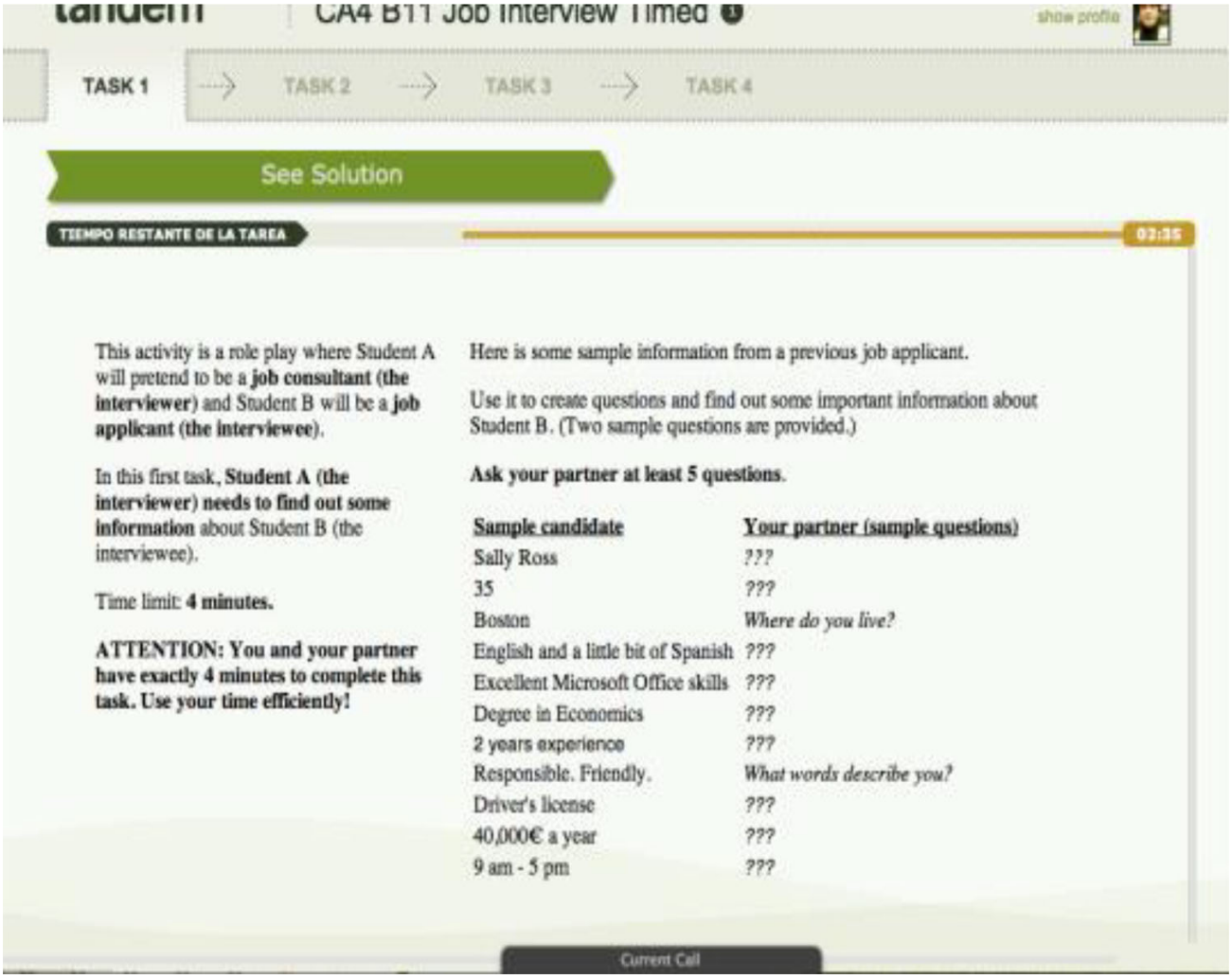


Figure 1. Screenshot of Task 1 in the role of interviewer.

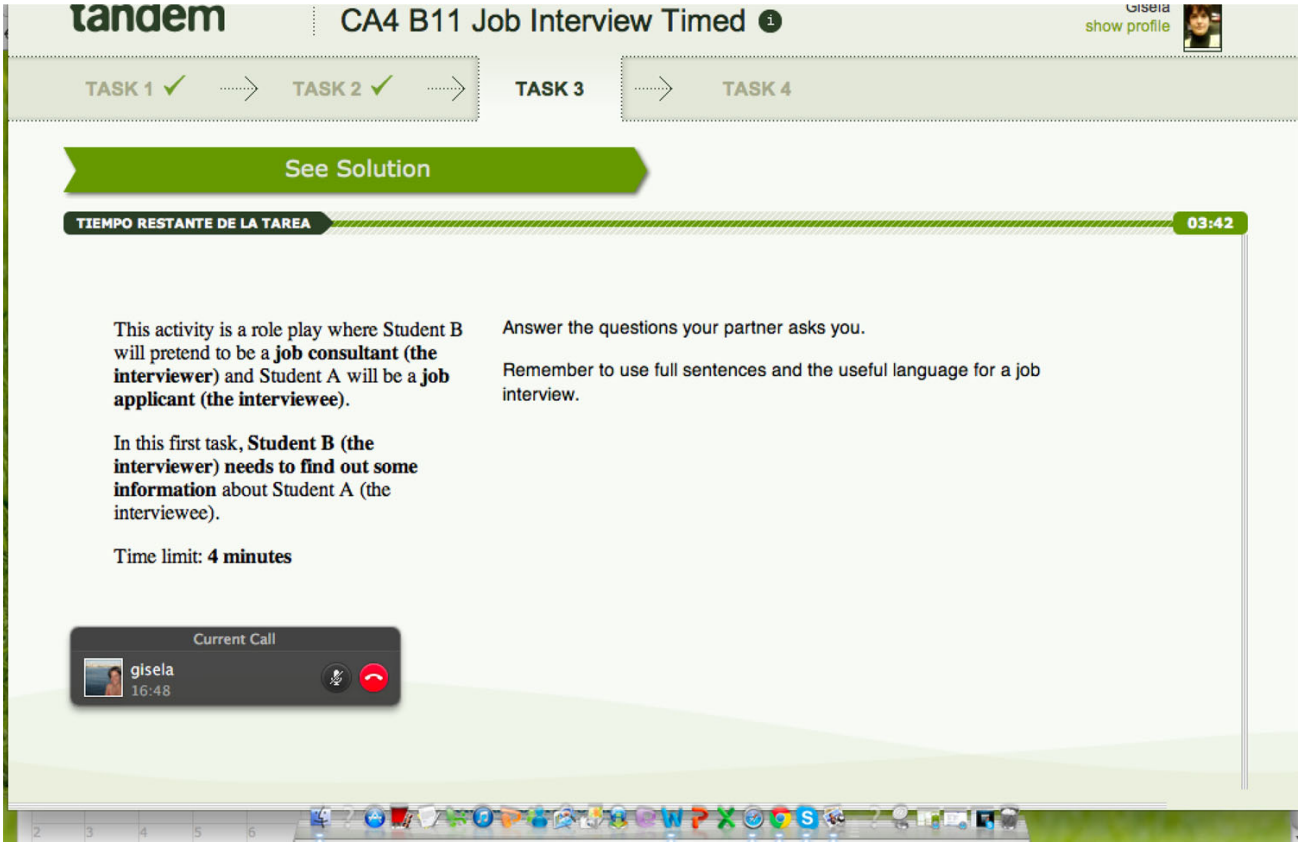


Figure 2. Screenshot of Task 4 in the role of interviewee. Note that Task 4 was the same as Task 1 but the roles were reversed.

between what learners were saying and the natural progression of the task or not. To answer research question 1, where words or phrases in the transcripts exactly or partially matched (e.g. 'agent' in text becomes 'agency' in spoken language), the resources in the screenshots were then highlighted in bold text in the transcripts. We identified and noted if and how each of dyad was encountering and using the text in some way and compared this across the dyads. In the analysis, we looked for evidence as to when and how learners were using screen-based resources in their speech. In this process, we developed the categories iteratively and could then identify what learners did with and/or how they used screen-based resources in relation to the content of their turns (the 'what' of the turns). The categories were: speech for conducting oral turns between learners, screen-based text becoming embedded in oral turns, screen-based text/visual prompting oral turns, screen-based visual/textual (navigational and pop-up) resources becoming topics of talk. We also looked at how learners orientated (following CA principles) to screen-based resources (the 'what/who' of the turn-taking). However, CA transcript conventions were not followed. Only the first two tasks were analysed, focusing on one learner as initiator/responder in the two tasks presented. The rationale for only analysing the first two tasks was that this would give us sufficient data for an exploratory study and tasks 3 and 4 were the same tasks repeated but with the students' roles reversed.

The data in this current study was analysed in a previous study by Knight, Dooly, and Barberà (2018). However, the research focus was very different. Knight et al. (2018) study explored the meaning making potential of navigational resources beyond the lingual. The main focus was on learners' physical negotiation with the technological tool and screen-based resources as well as the implications of the physical negotiation with the tool for online language learning as a social practice. In contrast, this current micro-level study is focused on learners' oral turns and how all the screen based resources, not just navigational resources, shape learners' oral turns in various ways.

6. Results and discussion

To answer RQ1 we identified that learners encountered and mediated with speech as a mode (spoken language as a resource) and other screen-based modes and resources throughout the task. This shaped the spoken turns and turn-taking between the dyads in different ways.

In relation to speech as mode used for interaction with their peer, an initiation/response sequence is evident through learners' verbal turns, many in the form of questions and answers e.g. in example 1 (lines 1

and 2), example 2 (lines 4 and 5) and example 3 (lines 10 and 11). Questions and answers were the most frequent turns and this is perhaps not surprising given that the task is to carry out an interview where question/answer pairs are the social norm. All names have been anonymised. Other discourse features included extended periods of silence (example 2 and 3), internal speech (example 2 and 3) and pauses and hesitations (examples 1, 2 and 3).

Example 1 (case 2): question and answer

1	N	Hello, good afternoon, (pause) how are you?
2	G	I'm fine, and you?
3	N	fine thank you

Example 2 (case 1): question and answer

		SILENCE and rustling of paper
1	L	(after 11 seconds) Ok paper SILENCE and rustling of paper
2	S	(after 23 seconds) Hi Laia
3	L	Hi
4	S	How are you today?
5	L	I'm fine thanks
6	S	Erm Sorry, what is your name?
7	L	My name is Laia, Laia Baños
8	S	Okay …
9	L	Yeah
10	S	sorry … how old are you?
11	L	I'm 34 years old
12	S	Er Well, Where do you live?
13	L	I'm live … I live.. I'm living near *of* Girona.

Example 3 (case 3): question and answer

		SILENCE
1	L	(after 18 seconds) Mhu
2	A	(Unintelligible) (PAUSE). I'm waiting. (PAUSE)
3	L	Well
4	A	Okay, (PAUSE) A or two?
5	L	(PAUSE) I have to answer your questions.
6	A	Okay, perfect.
7	L	So I'm student B and you are student A
8	A	Okay, perfect. I start. Well, thank you to be here, to came to this job interview; but I need some information about you. First, what is your name?
9	L	My name is Laia, Laia Sierra.
10	A	Okay, How old are you?
11	L	I'm a thirty-five, sorry thirty-seven years old.

Although learners began their talk somewhat differently they all orientated to each other as interviewer and interviewee with the question and answer turns dominating the talk. Learners had been previously assigned roles as interviewer (student A) and interviewee (student B), which had been communicated to them textually through the screen. This assignment of roles was designed to shape which learners carried out verbal turns of questions/answers as interviewer/interviewee as first or second position (talker) in the turn-taking.

In addition, based on the analysis of recordings, alongside the screenshots, there also appeared to be instances where learners encountered, orientated to and used the resources in different ways, which we will subsequently demonstrate. This also shaped the talk in different ways. The resources became relevant in their talk in that text from the screen became a) embedded or modified in oral turns; b) resources to initiate and support oral turns and c) topics of talk.

6.1. Embedded in learners' verbal turns

One way that learners orientated towards some resources was by embedding them as part of their talk. Example 4 is an excerpt from case 3 and is at the beginning of the recording. Student B cannot hear student A until they have clicked on 'accept' to the invitation initiated technologically by student A.

Example 4 (case 3):

		SILENCE
1	L	(after 18 seconds) Mhu
2	A	(Unintelligible) (PAUSE). **I'm waiting.** (PAUSE)
3	L	Well
4	A	Okay, (PAUSE) A or two?
5	L	(PAUSE) I have to answer your questions.
6	A	Okay, perfect.
7	L	So I'm **student B** and you are **student A**

After 18 seconds of waiting, in line 2, the learner (interviewer as student B) who must have invited learner (interviewee as student A) using the technological tool *Tandem*, used the phrase 'I'm waiting'. The turn of his peer is not a response to the previous turn of 'I'm waiting' because she could not have heard that turn unless she had accepted his invite before he uttered it (as with the tool in Liddicoat's study, 2010). In order for them to actually connect, the interaction must be both 'technologically and interactionally accomplished' (Tudini, 2014, p. 3). The invite must be carried out *kinaesthetically* by pressing the button.

The verbal turn 'I'm waiting' which follows 18 seconds later, indicates that while a physical response through the technology has been achieved (because the partner mumbles 'Mhu', line 1), the partner has not yet made an oral turn towards her partner. The computer generates the 'waiting' button after the student has initiated a turn, technologically, through the computer inviting his/her partner to join the conversation. The word 'waiting' in line 2 may have originated from the pop-up on the screen as seen in Figure 3 that also reads 'waiting' but we cannot confirm this. We hypothesise that the learner may have said 'I'm waiting' in order to update the teacher who will evaluate their recorded interaction or to inform his partner that he is ready to start interacting. The partner's mumbling may be a form of self-talk because she is orientating herself towards the screen (lines 3 and 5). Raudaskoski (1999) also found that the moment when a learner encountered text was 'muttering the words to himself' (p. 115). This result demonstrates that learners must attend to some screen-based resources to move forward in task completion. Unlike Yamada's (2009) finding that there is 'no image', results highlight the existence of many images (and modes) that learners must attend to as they 'multitask' (Knight & Barberà, 2016).

This result also revealed two things: the act of one learner inviting his partner using the 'start' button showed how learners can carry out turns *kinaesthetically*, echoing Liddicoat's (2010) finding that technology can be a co-constituent of the interaction. The second is the possibility that the screen-based text (the word 'waiting' on the screen) was being read aloud as a form of internal speech while waiting for his peer's technological turn of accepting the technological invite. This would mean that the screen-based word was being re-semiotized by the learner orally. We return to this potential phenomenon further on.

The textual information about the assignment of roles also appeared to embed itself in one of the learner's turns when the invited learner helped to orientate her partner by confirming information about their roles: 'So I'm student B and you are student A' as shown in Figure 4, line 7. This textual and oral information is used by the learner to help orientate her partner towards his role. Screen-text becoming embedded in oral turns also occurs with other screen-based resources which we subsequently present. We understand this use of resource (embedding text into oral turns) as learners orientating to 'others as signs' which then becomes part of the oral discourse. The content of learners' oral turns are shaped by multiple modes- through the integration of textual and visual signs in oral turns and thereby 'formed' transmodally. This echoes Lamy's (2006) finding that written/oral human turns were managed transmodally.

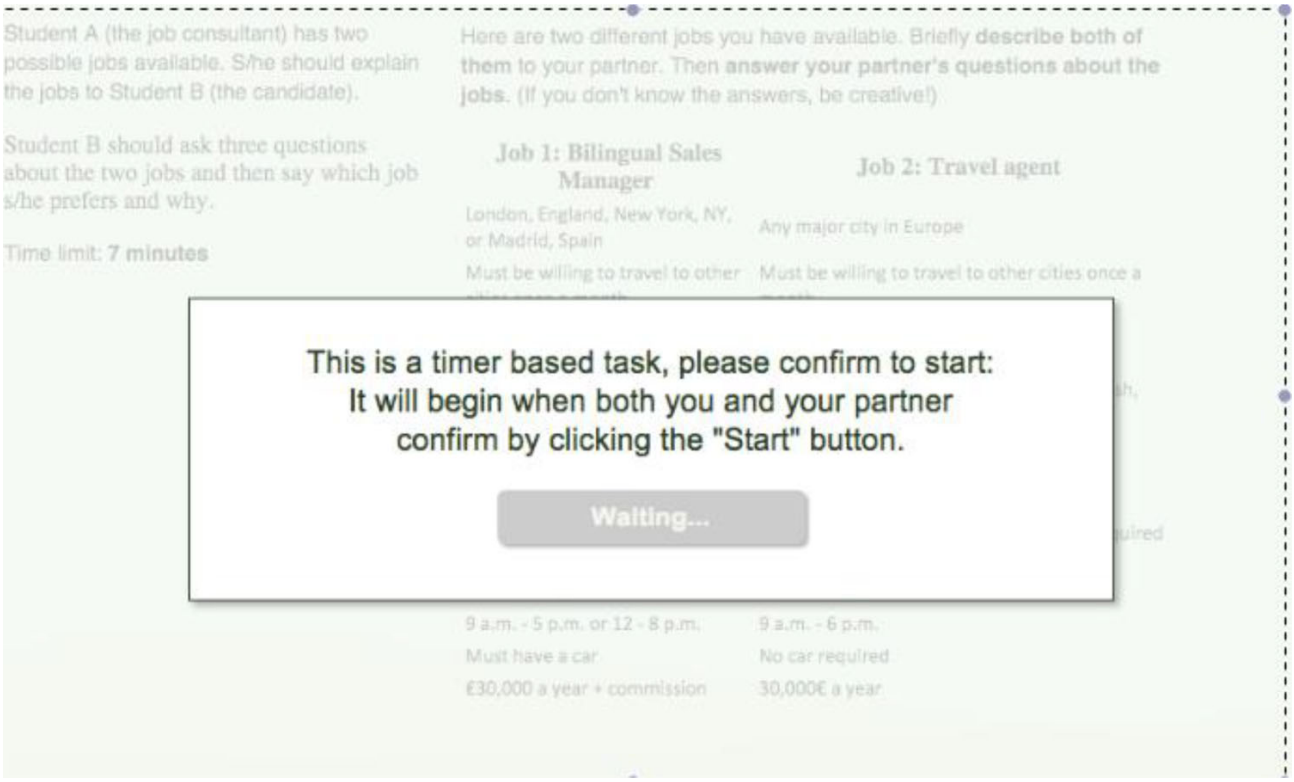


Figure 3. Screenshot at start of task with the text 'Waiting . . . '.

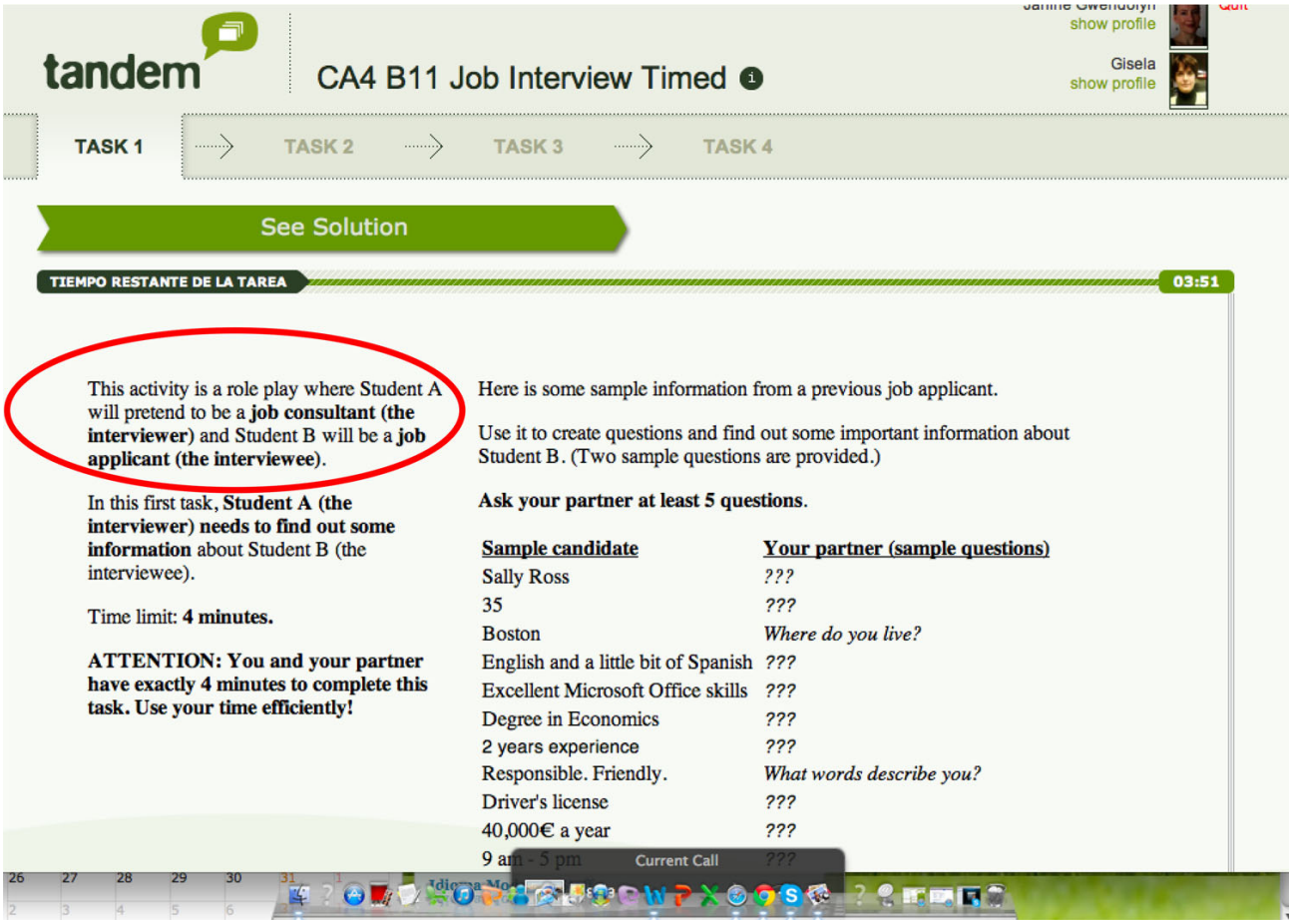


Figure 4. Textual information about students' roles.

6.2. Resources to initiate and form oral turns

Many learners orientated towards other textual resources (interviewers in tasks 1 and 2) as a support for question formation. Learners

(interviewers) asked their partners questions with many learners following a common order from the screen e.g. name, age, where they lived. Checking with the screenshot, the questions frequently asked by cases and the order they are asked, generally corresponded with the screen-based textual resources. Specifically, learners responded to the combination of resources which were: the textual request to act ('Ask your partner at least 5 questions'), the eleven textual resources ('sample candidate' and 'sample questions') combined with the sign '?'. Learners did not create questions very different from the text sample given, although some interviewers changed the question order or omitted questions in their turn-taking. Whilst the textual request to act can be understood as 'other as discourse' (speech act), the other textual resources for asking questions provide a supporting role for learners in the forming of oral turns.

Examples 5 and 6 show how these resources were used by the interviewer in Task 1 to support help him/her form questions. The bold and underlined text in the transcript shows how the questions correspond to the resources (the sign '?' and the sample answer) in the (edited) screenshot (Figure 5). The relationship of these resources highlights the importance of the 'interplay of modes' (Schnotz, 1999) pertaining to the screen-based resources, such as proximity to each other as a way of connecting them for learners, in the meaning making process.

Example 5 (case 1):

1 M	Hi
2 O	How are you today?
3 M	I'm fine thanks
4 O	Erm Sorry, **what is your name?**
5 M	My name is Maura, Maura Bater
6 O	Okay …
7 M	Yeah
8 O	sorry … **how old are you?**
9 M	I'm 34 years old
10 O	Er Well, **where do you live?**
11 M	I'm live … I live.. I'm living near *of* Girona.

Example 6 (case 3):

1 A	Okay, perfect. I start. Well, thank you to be here, to came to this job interview; but I need some information about you. First, **what is your name?**
2 L	My name is Laia, Laia Baños.
3 A	Okay, **How old are you?**
4 L	I'm a thirty-five, sorry thirty-seven years old.
5 A	**Where do you life?** Okay, (laughter) me too, no (laughter)
6 L	I live in Bellver de Cerdanya. In a little town in the Pyrenees.

All the participants in the cases used the question 'Where do you live?' as shown in Figure 5. It can be argued that this textual resource becomes embedded in the talk. However, not all cases used the second

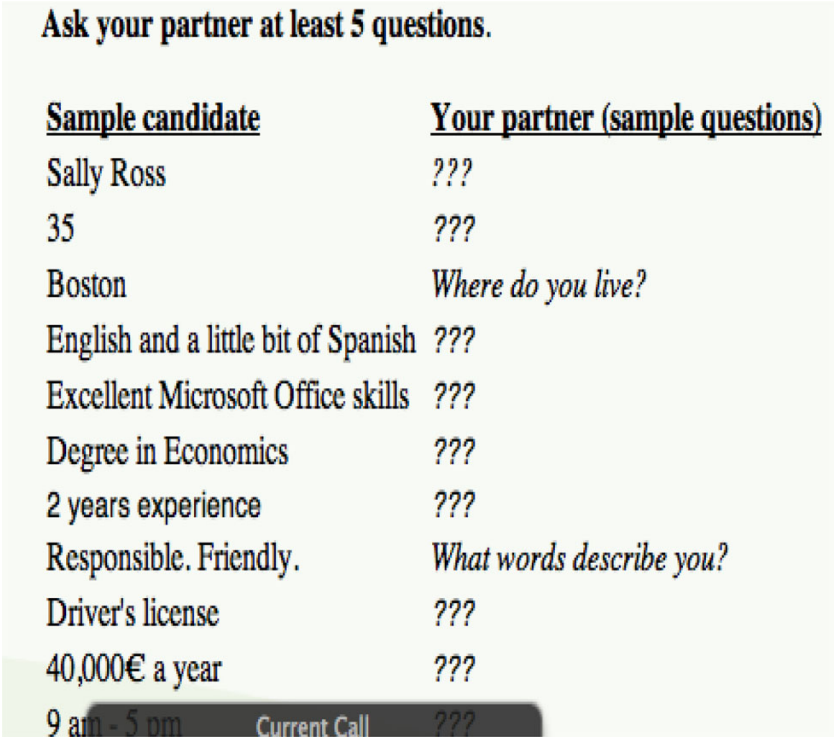


Figure 5. Resources for asking questions.

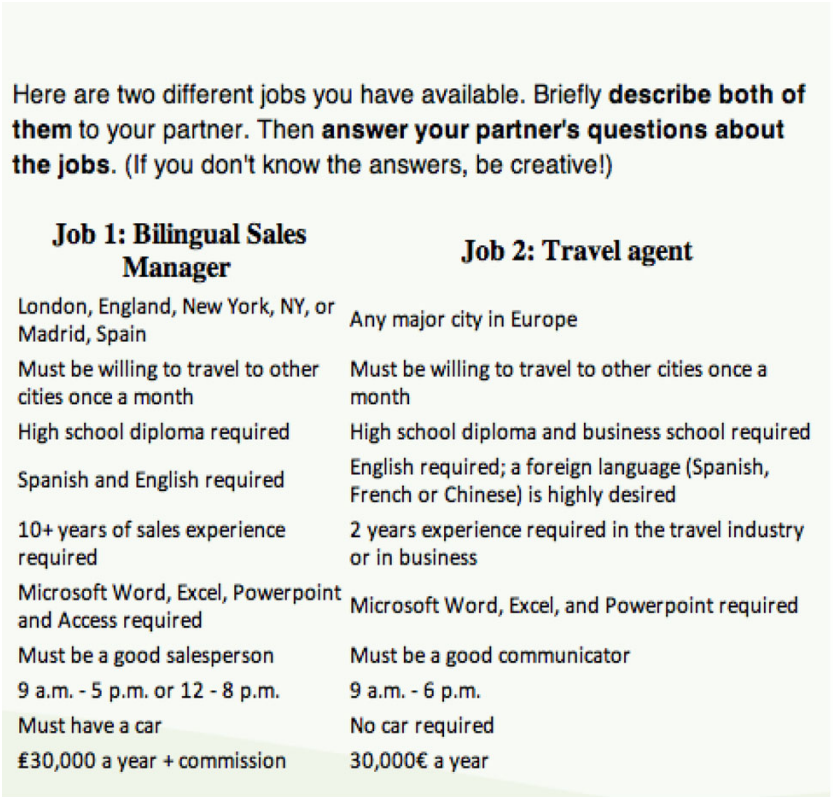


Figure 6. Textual information for describing two jobs.

complete sentence given: 'What words describe you?' highlighting that learners' choices over how they chose to orientate to textual scaffolds may change from moment to moment. The questions 'what is your name? and 'how old are you?' are uttered by the interviewers in most cases. However, these questions do not exist as text on the screen. The symbol '?' and suggested answer '35′ (as shown in Figure 5 above)

provide support for the learner's turn, which we understand as 'others as signs'. Learners also modify the text so that for example, 'city in Europe' in the text (Figure 6 in the 7th line down under 'Job 2: Travel Agent') becomes 'European city', spoken, in example 7, case 2, line 10.

Similar to task 1, in task 2 there are differences in the ways in which the learners orientate towards the screen-based text. In example 7, the interviewer in case 2 uses the two jobs to help structure her talk (lines 1, 3, 5 and 7). She uses the textual information (highlighted in bold in Figure 8) to describe the jobs to her partner and also to help her respond to her partner's questions about the job. Rather than the text initiating or supporting oral turns, the text from the job descriptions become embedded in her talk as we saw previously with the word 'waiting' on the navigational resource in Figure 3.

Example 7 (case 2):

1G	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. Do you have any questions about these jobs?
2N	Yes … what is the time table?
3G	Ok. In the bilingual sales managers is two times to choose: in the morning you work in the nine am, until five pm, and yes in the afternoon you can go from twelve o'clock until eight pm in the afternoon.
4N	Ok and what is the salary?
5G	Is very similar for both jobs: in the bilingual sales manager is thirty pounds thirty thousand pounds sorry a year but there are commissions, if you sales (sell) more you earn more, of course. And the travel agent there are thirty thousands a year too, but is a salary fixed no by commission.
6N	Ok, and about first job, I should to travel every day or every week?
7G	No, only you must be travel once a month and go to cities … big cities like London, New York and Madrid once a month, because English and Spanish is very important speak fluently.

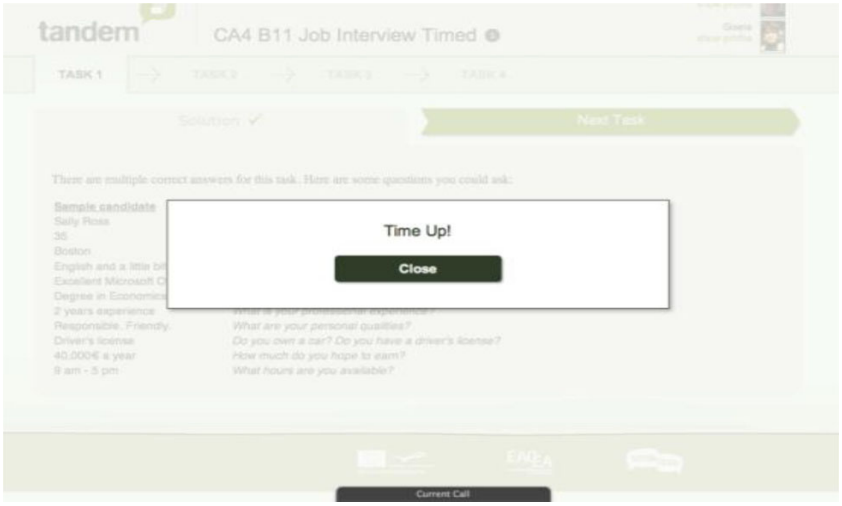


Figure 7. Time Up! Pop-up.

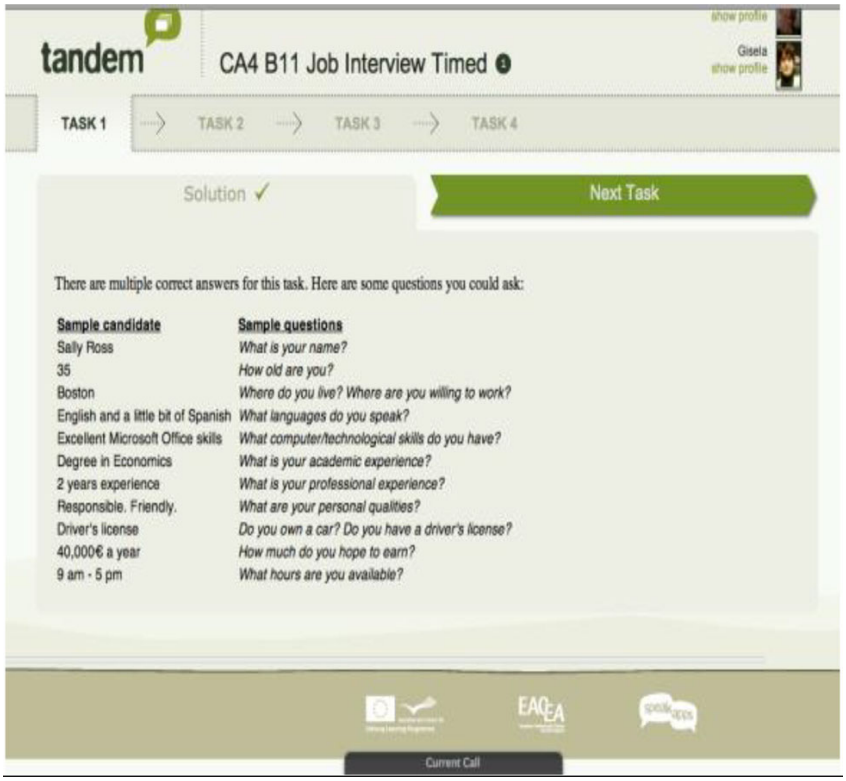


Figure 8. Solutions page.

In contrast, in example 8, we can see how the interviewer in case 4, does not use the textual job descriptions given and this shapes their talk.

Example 8 (case 4):

1	T	And what is your professional experience (unintelligible)
2	M	Errrr eight years experience with students six to twelve years old
3	T	What are your personal qualities?
4	M	I'm very responsibly and (PAUSE) I think that I'm friendly and I have a lot of patience and I'm a sociable person.
5	T	How much do you hope to earn?
6	M	I can work for less one thousand euros.
7	T	(PAUSE) thank you Mona, I can offer you two different jobs in my school. For one hand I have a place in primary with students six to twelve years old, for the other hand I have a place in secondary with students twelve to sixteen years [old](yes, Mona interrupts) in two years you will be in this teacher (Mona interrupts)
8	M	[yes]
9	T	in two years you will be in this teacher [what do you prefer?]
10	M	[Mmmmmm Montse (pseudonym of student)] I have a question before I decide .... Mmmm which work time table?
11	T	(PAUSE) well ... in primary school you will work at night o'clock to half past twelve. On the secondary school you'll work at eight o'clock to five o'clock PM
12	M	Mmm and ... how many groups of students will I have?
13	T	(PAUSE) in primary school you will have three groups ... of students ... (cough) Sorry ... and in secondary school eight per group.
14	M	mmm and one question ... how much will I earn?
15	T	(PAUSE) you will earn eight hundred euros in primary school and in secondary school your salary will be two thousand euros (Mona makes humming noises)
16	M	Mmm for me two jobs are very interesting but I prefer to work ... to be an English teacher in secondary school because I haven't any experience high school and I would like to work in secondary school ...
17	T	(PAUSE) Oh very good, Mona. Is fantastic. The job is for you! I hope you next Monday at 9 o' clock in my school

Following task design, at this point of the task the screen-based text (job descriptions pertaining to a bi-lingual sales manager or travel agent) should be the same as case 2, in example 7. However, what becomes apparent in case 4's talk is that they did not use the job descriptions provided. This omission shaped the interaction. Instead, the interviewer asserted (previous to the extract presented) that 'I'm a headmistress in a private school in Barcelona' and asked questions related to this job description. However, there is some evidence that the interviewer may have used the textual resources from the previous screen (Figure 5) to form questions because she refers to 'experience' (line 2) and 'how much do you hope to earn?' (line 5), which indicates that they were not following the interface pages of the Tandem tool in the way it was designed i.e. looking at specific interface pages at the moment of talking.

On the one hand, the choice not to use the screen-based job descriptions afforded the learners some creativity, as they created their own jobs. On the other hand, what starts to emerge in the turns and turn-taking is that the interviewee begins to ask questions to the interviewer. The roles of who has 'first position' as interviewer in the turn-taking process and who is the responder (roles of A and B) become blurred. The interviewee orientates to her partner as the questioner (line 10), contravening her pre-given role. This result suggests that the learners' roles of A and B (interviewer and interviewee presented to learners on a previous screen), were not being adhered to. The interviewee 'took the floor' (beyond her role) assuming the role of questioner rather than responder (lines 12 and 14). Therefore, screen-based resources (the communication of roles: who is A and B) may act as a regulatory and/or guiding presence in the absence of a teacher during tasks. Learners' choices to modify their given roles (communicated through different screens) may support possibilities for greater creativity between learners but may also allow for role reversal or allow for oral turns (and time talking) to be less equitable.

The results echo Raudaskoski's (1999) finding that the material/visual semiotics can become 'embedded in the sequential meaning making' (Raudaskoski, 1999, p. 132). They appear to initiate learners to act orally (task instruction) and also are orientated towards as support for turn formation. Learners can orientate to some textual/visual resources in a way that the making of verbal turns and meaning making becomes co-shared between the text and the talker, which then enters the discourse. Learners can also choose not to use some textual/visual resources in their negotiation of verbal turns and this may have both positive and negative consequences for task-based talk.

6.3. Form part of the topic of talk

On occasions the resources become a topic of the talk between learners. In example 9 below, the dyads have come to the end of Task 1 and the interviewer has asked the questions to the interviewee. The excerpt starts with a response from the interviewee to how much she hopes to earn (line 1).

Example 9 (case 3):

1 L	Oh, well … I have not great aspiration and perhaps a thirty thousand … Euros per year.
2 A	Okay, perfect! Okay, thank you. I look for the position we have. Okay … well ok..now … **close**, no, I suppose … ?
3 L	I suppose too, I **close** it …
4 A	Well here are the **questions** and the **answers** … We go to the **next task** … ?
5 L	Sorry?
6 A	We go to the **next task**?
7 L	Okay, huh … But the computer say me that the **time** **is** **up**.
8 A	Ah, yes.
9 L	So, I think, I think, Tandem is not working correctly.
10 A	No, maybe no because I'm in another page, is in the **solutions page.**

As the learners neared the end of the pedagogical task, learner A says in line 2, 'Okay … well ok.now … **close**, no, I suppose … ?' to which his partner responded in line 3, 'I suppose too, I **close** it … ' Because the mention of closing is a new topic of talk, checking the screenshot gives insight into what the meaning may be. It becomes evident that the learners were referring to the Time Up! pop-up which has the text 'close' on it as shown in Figure 7 above. The word 'close' became a topic of talk as both learners engage in a brief discussion as to whether they should close the 'Time Up!' pop-up or not. While the word 'close' can be interpreted as an 'invite' or 'other as discourse/speech act' the learner orientated to it as an instruction to act that needs to be negotiated orally with their partner with respect to how they might respond ('**close**, no, I suppose … ?' in line 2). They would need to click individually (physically) but negotiate jointly (orally) so that both learners click simultaneously in order to be on the same page. This is possibly made a more difficult task by the fact that they cannot see each other's gestures or gaze.

Both learners used the word 'close' orally (line 2 and 3) as a verb and one learner used the verb 'go' with 'Next Task' (lines 4 and 6). This suggests that learners orientated to the resources as something that they needed to deal with physically, echoing the *procedural* message type (furthering task completion) identified by Kraut et al. (2003). The 'Next Task' button invited them both to respond physically and they negotiated orally as to whether they are going to carry out that physical (or *kinaesthetic*) response as a turn. The navigational resources become 'third

others', made relevant in the talk by the learners when they needed to act simultaneously. The need for learners to multi-task during activities (Knight & Barberà, 2016) with respect to carrying out navigational moves, while attending to the pedagogical tasks becomes evident.

This is followed by the partner's reference to the Time Up! pop up screen resource as an 'it' (line 3). The learner is explicit referring to 'it' having orientated to the resource as 'a third other' that is requesting him to 'close'. By clicking on 'close' we understand that he would be accepting that invite technologically. This echoes Raudaskoski's (1999) finding of how learners referred to the computer/program using the term 'this one'. It is 'ambiguous in human/inhuman respect, but which, nevertheless objectifies the 'other' into a non-copresent entity' (Raudaskoski, 1999, p. 21). The resource is orientated to as 'other as social entity'. The reference to 'it' (line 3) also reflects Kraut et al. (2003) *referential* message type (identification/location of task objects).

Following on from the oral negotiation of a navigational act, student A in line 4, indicated that he was on the page with questions and answers (which we hypothesise to be the solutions page, Figure 8). He asked his partner whether they should go to the next task. The 'Next Task' button is also on the solutions page and is a textual/visual/navigational resource. However, it becomes clear with his partner's 'Sorry?' that she was not saying sorry because she did not hear him, but rather because she did not understand what he was talking about. She says that 'The computer say me the time is up' (line 7), indicating that she was not seeing the same resources (therefore nor the same interface) because she was on 'another page' (line 10). Ultimately, she decided that the tool was not working correctly. The learner's comment 'Sorry?' (line 5) also pertains to Kraut et al. (2003) *internal state* message type (relating to intentions, knowledge, emotions etc.).

Apart from the resources ('Close' and 'Next Task') that learners must use to navigate, both learners also made reference to the resources in order to communicate and check what page they were looking at on the screen. The use of deixis ('here') emphasised location and a need to highlight learners' proximity in relation to a screen-based resource, another *referential* message from Kraut, Fussell and Siegel's message types (2003). 'Well here are the questions and answers' (line 4) and later his partner said 'I'm in another page, is in the solutions page' (line 10). The words 'here' (line 4) and 'I'm in' (line 10) emphasise learners' apparent need to establish and check what their partner is seeing. Their talk becomes a tool for physical/visual orientation for them. Their use of verbal summary of what is on the screen ('here are the questions and answers' and 'in the solutions page') rather than direct referencing of the

resources on the screen, appears to suggest that once learners realize they are not on the same page or have 'lost' each other, the mediation of smaller screen-based resources, which were important in the pedagogical task, are no longer useful. Their lens for mediating to problem solve has expanded outwards to include mediation through verbal references with the interface 'pages' as a means to share mutual understanding about their problem. We liken this to the expanding aperture of a camera lens. The multimodal ensemble is not so important at this moment but rather the interface page itself becomes the important resource.

Thus, in relation to research question 2, regarding how learners initiate or respond to other screen-based resources (and what this tells us about turn-taking), we identified that learners respond to other screen-based resources orally and *kinaesthetically* which is made relevant in their oral (and physical) turns. Screen-based resources also appear to initiate turns (task instruction to 'ask questions') and navigational resources can become a topic of talk. In addition, learners must 'initiate' with some resources physically ('start' and 'next task'). These can be considered passive agents because they respond to a learner's initiation (to move learners to another place) through touch (the *kinaesthetic* mode). Learners may also need to 'accept' as a response to navigational resources such as pop-ups.

We now turn to the pop-up resource which we identify as an 'active agent' (Dourish, Bellotti, Mackay, & Ma, 1993) and turn-taker in the discourse, adding to our understanding of how learners respond to pop-ups as 'others as (active) agents'.

6.4. The pop-up: the interrupter and the messenger

Not only did the text from navigational resources appear to become embedded in learner talk ('waiting', example 4, case 3) or become a topic of talk ('close' on the pop-up and 'next task' button in example 9, case 3) but learners orientated to them in other ways. The first is with regard to the Time-up pop-up daemon (Figure 7), which is orientated towards by learners as an interrupter of a turn and secondly, as a deliverer of a message. Both ways highlight that learners can orientate towards navigational resources as 'others as social entities' or agents.

6.4.1. The interrupter. In example 10 below, the learners were coming to the end of Task 2 and the interviewee is clarifying some aspects of the two jobs that her partner, the interviewer, had presented her with. After declaring 'Yes, I prefer the second job' (line 7), she follows with 'Time Up! Ok' (line 7) and continues in the same turn with why she prefers

the second job. The learner orientated towards the pop-up as a brief interrupter of the conversation with her partner, which she noted/acknowledged ('ok' in line 7) and which the learners manage, calmly, in order to continue the closure of their interview.

Example 10 (case 1):

1	L	You have said I need a diploma or qualification, concrete qualification, *specially* (special) qualification? Really?
2	S	Yes, diploma, high school diploma. And business school
3	L	Business school … okay, okay. Well without commission … and the commissions … do you *have* about what commissions can be?
4	S	I'm sorry I don't, I don't know.
5	L	Okay, well …
6	S	Do you prefer?
7	L	Yes, I think I prefer the second job. **Time up!** Ok, well I prefer the second job but I don't know if I can choice that because I don't have the experiences in industries but for me it's better, okay?
8	S	Okay good, thank you very much for your time, it's a pleasure er meeting you today (interference).
9	L	Okay me too thanks

6.4.2. The messenger. In another case, a learner orientated to the same resource differently, as shown in example 9, case 3 previously. The learner said 'Okay, huh … But the computer say me that the time is up'. The 'other' here (this time referred to as 'the computer') is referred to like a messenger. The 'message' only appeared on her screen (following task design) which she then conveyed to her partner as an explanation as to why she could not fulfil her partner's request that they move to the next task.

In relation to facilitating spoken interaction and therefore L2 development, the synchronous CMC tool Speak App itself was found to facilitate spoken interaction *through* the screen, just like other CMC synchronous audio and video tecnologies noted by Blake (2016). In addition, the results also suggest that the different multimodal resources *on* the screen also had an important part to play in facilitating interaction. Specifically, these resources provided the task topic, assigned interlocutor roles, provided lexical items and structures for learners' talk, specified the amount of turns initiated, prompted talk initiations and provided metalanguage in relation to the techno-pedagogical aspects of the task (such as words on pop-ups).

The screen, with the interplay of the various multimodal resources, was able to successfully represent the 'voice' of the language teacher and also provide the necessary instructional guidance related to student roles, scaffolding language for a final communicative goal and providing topic-related resources typically found in a task-based approach to language teaching.

7. Conclusions

This study has explored how the online nature of the participants' interaction was a relevant constituent part of learners' interactions (following

Liddicoat, 2010), not just a facilitative channel for interaction. The focus on semiotic resource-and-mode(s) rather than mode-as-channel has played an important role in highlighting how learners' various orientations to screen-based resources can lead to differences in turns and turn-taking. The resources shaped learners' talk by becoming embedded, modified in oral turns; initiating and supporting oral turns; becoming topics of talk as well as orientated towards as turn-takers.

The study contributes to SLA studies by highlighting that attending to the 'material stuff' (Kress, 2003) on the screen enables teachers and researchers to understand learners' interactions as 'a multimodal experience' (Collentine, 2009) more completely: not only understand the oral interactions between students beyond the pedagogical task but also the (inter)actions that learners' have with screen-based resources. This supports a shift away from 'the lingual bias' (Block, 2013) in SLA studies to understanding oral interaction more holistically. It can also help promote competent 'online communication' which 'is always mediated through a machine' so therefore 'unlikely ever to be exactly the same as face-to-face interaction' (COE, 2018, p. 96). As the CEFRL descriptors indicate, 'there are emergent properties of group interaction online that are almost impossible to capture in traditional competence scales'; including being able to mediate effectively 'the availability of resources shared in real time' (Ibid.).

Identifying how the screen-based resources has shaped learners' oral turns is useful for teachers because it highlights the different roles and purposes that screen-based resources can have in the design of oral interaction tasks. It can also help language teachers design online language learning tasks and assessment that is 'directed towards enabling learners to act in real-life situations, expressing themselves and accomplishing tasks of different natures'; that is, providing learners with 'communicative ability in real life' (COE, 2018, p. 27).

Teachers may wish to consider how to use the resources to encourage learners' use of specific lexical items, for example, or to scaffold the initiation and/or maintenance of learner talk. In addition, the study highlights how learners' use the meta language of the task relating to the technological aspects, such as 'time up' and 'close'. Implications for task design may be to attend to these by checking the words on these resources are language level appropriate and consistent throughout all future tasks. Cultural considerations should also be attended to. For example, the colours red, yellow and green are used to communicate meaning to learners about how much time is left, presumably following a traffic light system common in many European cities. However, in other countries, such as Japan for example, other colours are used to communicate the same meaning.

The results also highlighted tasks facilited L2 development because the screen and the multimodal resources replaced a number of roles and purposes needed for the students to accomplish the task successfully. The teacher role in particular, was replaced by different resources: the written text provided students with the topic area and lexical items as well as textual structures as scaffolding support for learners turns. Furthermore, the navigational resources and pop-ups supported task management with relation to time regulation and sequencing of different parts of the task. These are useful for teachers because they can see how typical pedagogical aspects of tasks can be orchestrated and managed through interface design and features.

Language teaching through audioconferencing (and potentially videoconferencing), as well as research into it, must take into account that students may use and learn the meta language related to the technology as well as the language related to the pedagogical task (e.g. start, close, next task) and this should be expected as a naturally occurring part of task talk.

Regarding research, the methodological implications may be of use for scholars who carry out analysis on oral learners' turn-taking and want to further explore how students' oral turns are shaped by their use of other screen-based resources during specific tasks.

Future research might focus on the effects of physical activity inclusing navigating and reading simultaneously on the learner during oral tasks such as effects on their cognitive load or whether some screen-based resources can be overly distracting during talk, for example. Such future research might look to frameworks in the HCI literature such as Kraut et al. (2003) when analysing spoken interaction, particularly the process of meaning negotiation.

In addition, future research on multimodal aspects of task design might focus on the interface design, including layout and colour, not just on the tool design. Furthermore, adjustments to whether different screen-based resources (providing more/less visual prompts or instructing learners' textually to ask 8 instead of 5 questions) supports an increased amount of spoken interaction would be particularly useful.

7.1. Limitations of the study

The insights gathered into how learners' oral meaning making is shaped through both verbal and nonverbal resources in the multimodal experience described in the paper is limited in depth and breadth. The pool of data analysed is limited, not longitudinal in nature and the study only

focuses on one type of interactive task. In addition, the analysis of learners' use of one digital tool is used.

Therefore, these limitations should be taken into account in any similar future studies with the same aim or that seek replication.

ORCID

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

References

Austin, J.L. (1962). *How to do things with words*. Cambridge: Harvard University Press.

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

Bezemer, J., & Kress, G. (2016). *Multimodality, learning and communication: A social semiotic frame*. London: Routledge.

Blake, R. (2016). Technology and the four skills. *Language Learning & Technology*, *20*(2), 129–142.

Block, D. (2013). Moving beyond lingualism: Multilingual embodiment and multimodality in SLA. In S. May (Ed.), *The multilingual turn: Implications for SLA, TESOL, and bilingual education* (pp. 54–77). New York: New York.

Calvo-Ferrer, J. R., Melchor-Couto, S., & Jauregi, K. (2016). ReCall special issue: Multimodal environments in CALL editorial multimodality in CALL. *ReCALL*, *28*(3), 247–252. doi:10.1017/S0958344016000136

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

Coulthard, M. (1985). *An introduction to discourse analysis* (2nd ed., 1st ed., 1977). London: Longman.

Council of Europe (COE). (2018). *Common European framework of reference for languages: Learning teaching, assessment companion volume with new descriptors*. Strasbourg: Council of Europe.

de la Colina, A. A., & García Mayo, M. D. P. (2007). Attention to form across collaborative tasks by low-proficiency learners in an EFL setting. In M. P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 91–116). Clevedon: Multilingual Matters.

Dourish, P., Bellotti, V., Mackay, W., & Ma, C. Y. (1993, December). Information and context: Lessons from the study of two shared information systems. In *Proceedings of the conference on Organizational computing systems* (pp. 42–51). New York, NY: ACM.

Gass, S. M., & Mackey, A. (2006). Input, interaction and output: An overview. *AILA Review*, *19*, 3–17. doi:10.1075/aila.19.03gas

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. doi:10.1017/S0267190515000112

González-Lloret, M. (2011). Conversation analysis of computer-mediated communication. *Calico Journal*, *28*(2), 308–325. doi:10.11139/cj.28.2.308-325

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts* (pp. 41–58). New York: Academic Press.

Guth, S., & Helm, F. (2012). Developing multiliteracies in ELT through telecollaboration. *ELT Journal*, *66*(1), 42–51. doi:10.1093/elt/ccr027

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

Jenks, C.J. (2014). *Social interaction in second language chat rooms*. Edinburgh: 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.), *Conference proceedings of the 9th international conference of ICT and language learning* (pp. 248–252). Italy: Librariauniversitaria.it edizioni.

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

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. doi:10.1017/S0958344019000132

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. doi:10.1207/S15327051HCI1812_2

Kress, G. (2003). *Literacy in the new media age*. London: Routledge.

Lamy, M.-N. (2006). Multimodality in second language conversations online: Looking for a methodology. In *proceedings of third international conference on multimodality* (pp. 385–403). Pavia, Italy.

Lantolf, J. (2000). Introducing sociocultural theory. In J. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 1–26). Oxford: Oxford University Press.

Levy, M., & Kennedy, C. (2004). A task-cycling pedagogy using audio-conferencing and stimulated reflection for foreign language learning. *Language Learning and Technology*, *8*(2), 50–68.

Liddicoat, A. J. (2010). Enacting participation: Hybrid modalities in on-line video conversation. In C. Develotte, R. Kern, & M. N. Lamy (Eds.), *Décrire la conversation enligne* (pp. 37–50). Lyon, France: ENS Editions.

Liou, H. (2011). Blogging, collaborative writing, and multimodal literacy in an EFL context. In M. Levy, F. Blin, C. B. Siskin, & O. Takeuchi (Eds.), *WorldCALL: International perspectives on computer-assisted language learning* (pp. 3–18). London: Routledge.

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

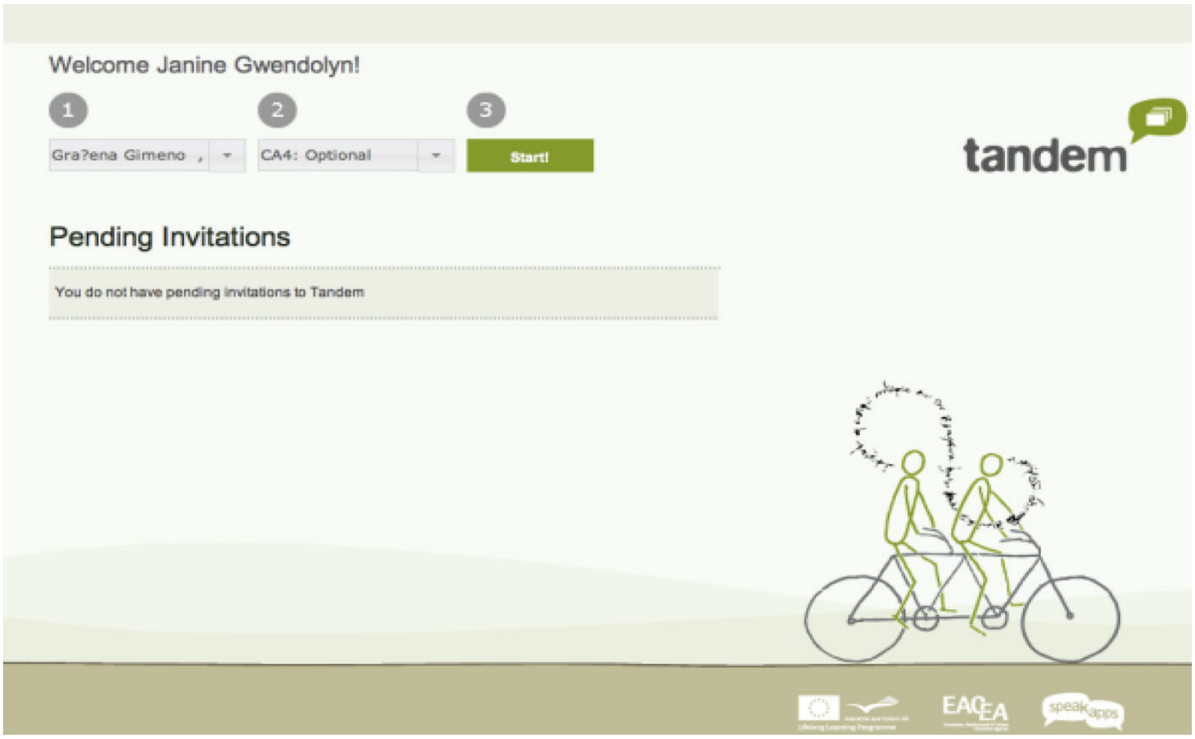
Sacks, H., Schegloff, E.A., & Jefferson, G. (1974). A simplest systematics for the organisation of turn-taking for conversation. *Language*, *50*(4), 696–735. doi:10.1353/lan.1974.0010

Schnotz, W. (1999). Visual learning with new technologies: Introduction. *European Journal of Psychology of Education*, *14*(2), 163–165. doi:10.1007/BF03172963

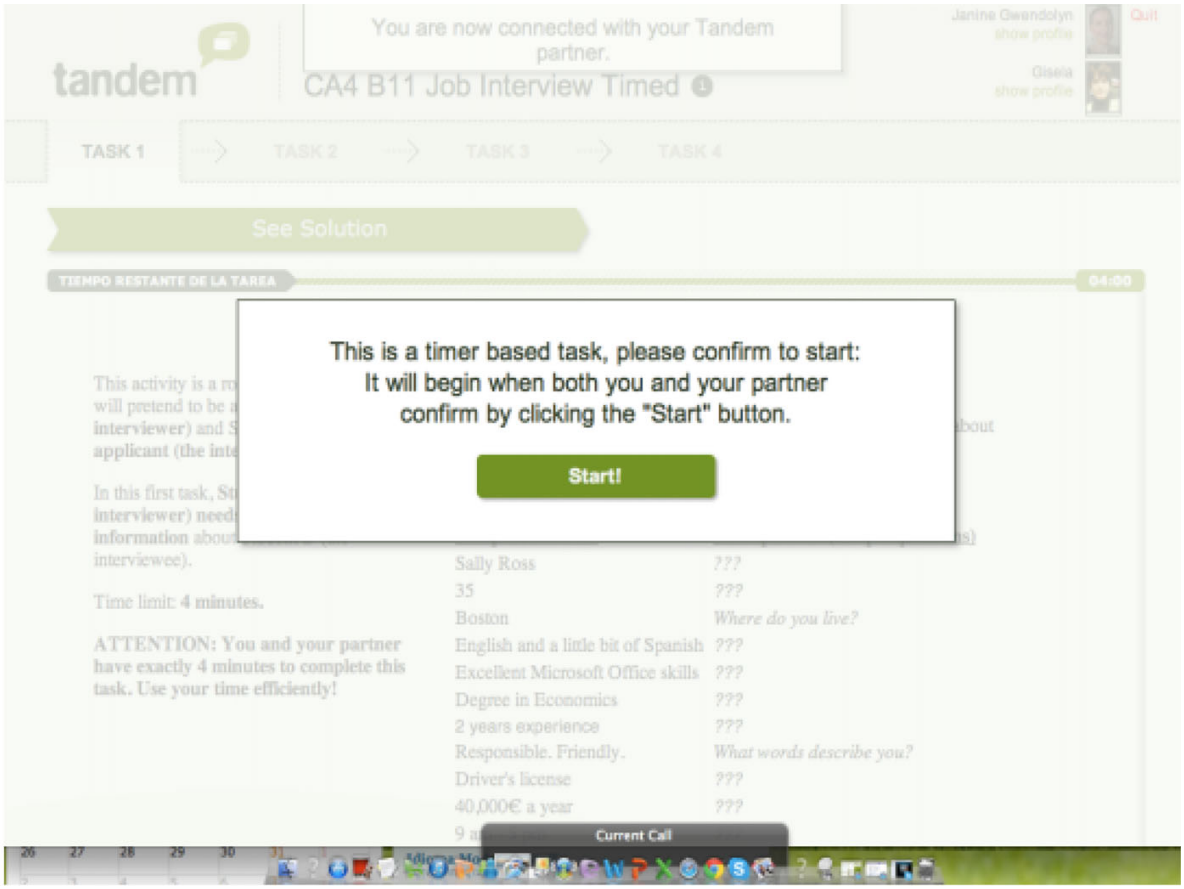
Searle, J. (1969). *Speech acts*. Cambridge: Cambridge University Press.
Sinclair, J. M., & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford: Oxford Univ Press.
Stenström, A. B. (1994). *An introduction to spoken interaction*. London: Longman.
Tudini, V. (2014). Conversation analysis of computer-mediated interactions. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–7). Hoboken, NJ: John Wiley & Sons.
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. doi:10.1016/j.compedu.2008.12.007

Appendix A: A sample of screenshots from the reconstructed task collected from researcher simulation

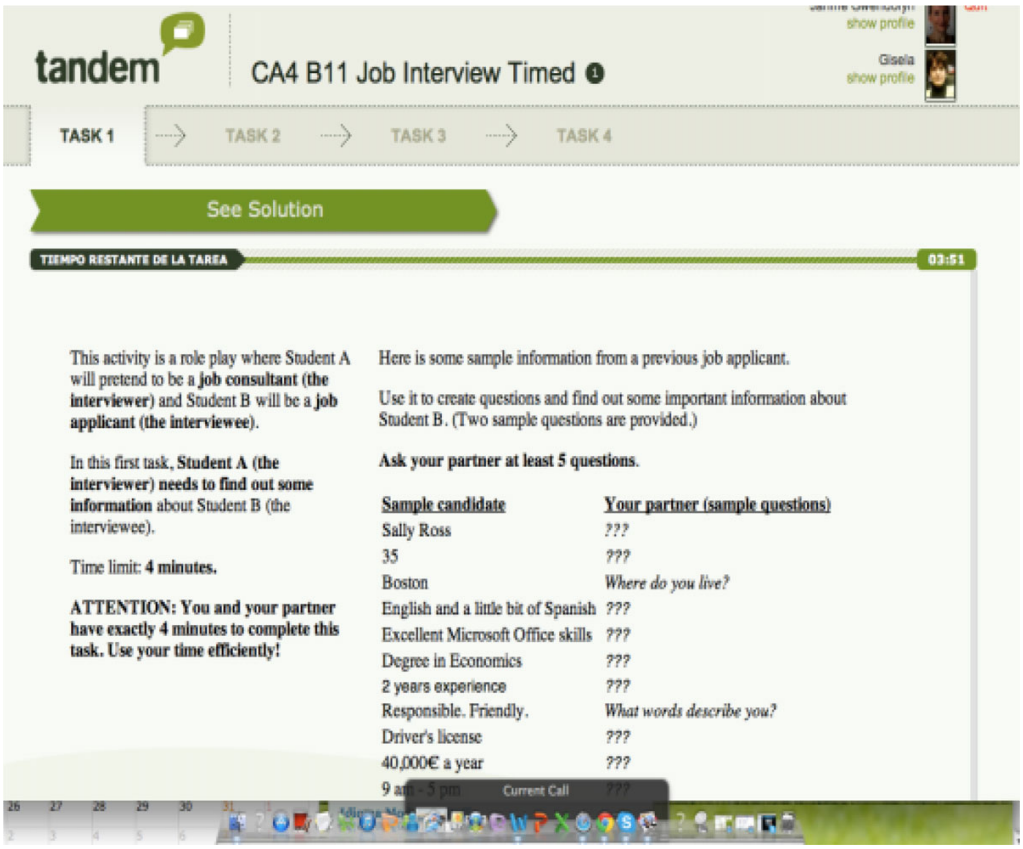
1.



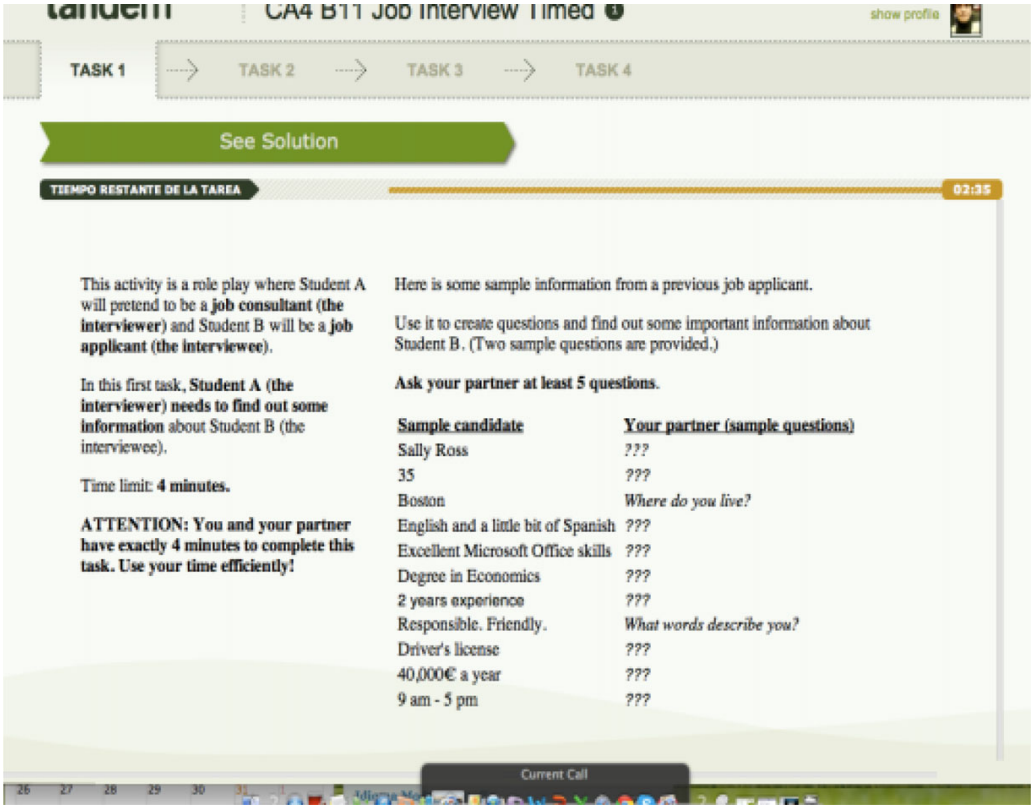
2.


You are now connected with your Tandem partner.
tandem
CA4 B11 Job Interview Timed
Janine Gwendolyn
show profile
Quit
Gisela
show profile
TASK 1
TASK 2
TASK 3
TASK 4
See Solution
TIEMPO RESTANTE DE LA TAREA
04:00
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.
Start!
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 ???
Current Call

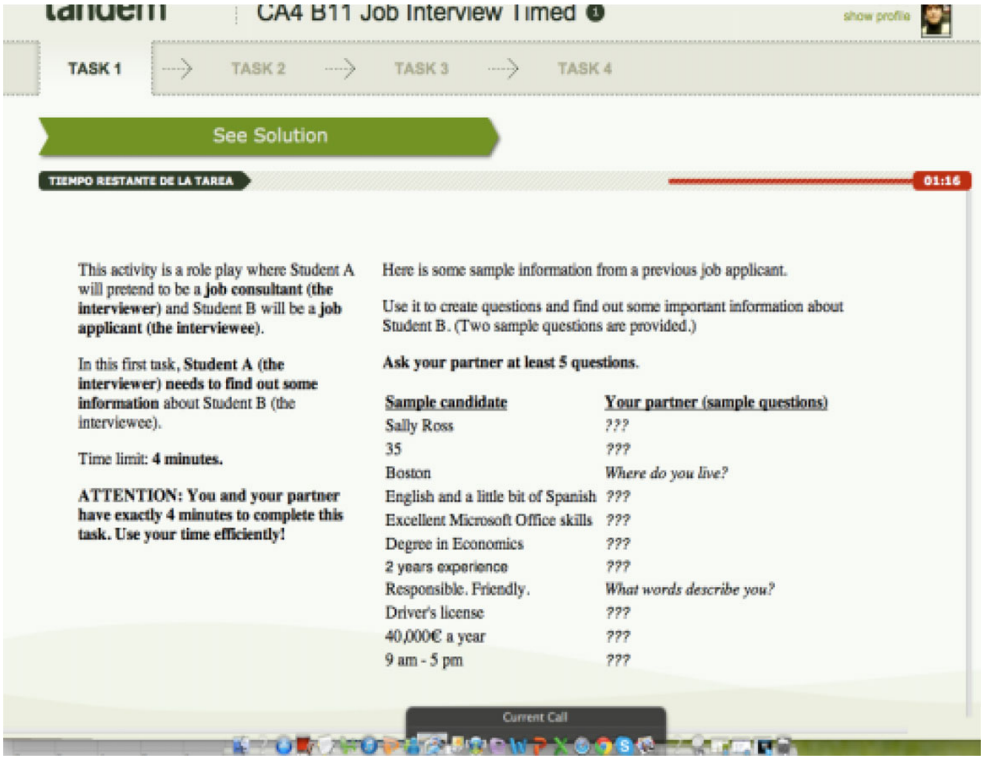
3.


tandem
CA4 B11 Job Interview Timed
show profile
Gisela
show profile
TASK 1
TASK 2
TASK 3
TASK 4
See Solution
TIEMPO RESTANTE DE LA TAREA
03:51
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 ???
Current Call

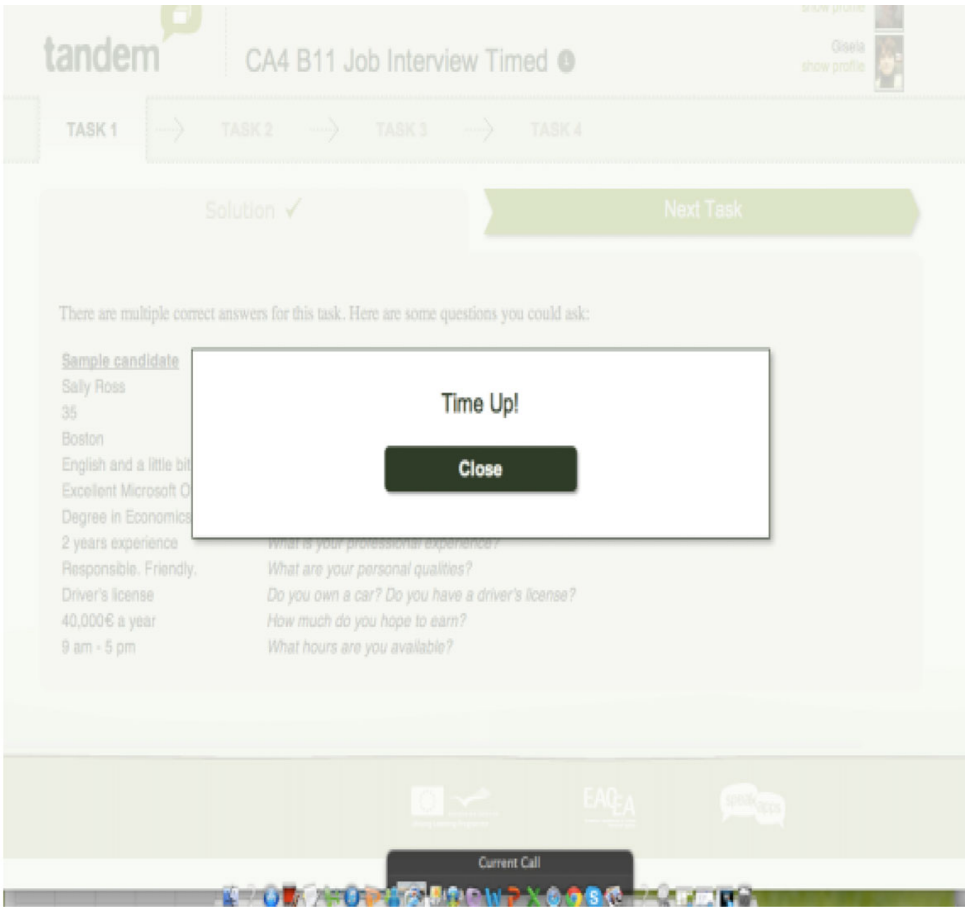
4.


tandem
CA4 B11 Job Interview Timed
show profile
TASK 1
TASK 2
TASK 3
TASK 4
See Solution
TIEMPO RESTANTE DE LA TAREA
02:35
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 ???
Current Call

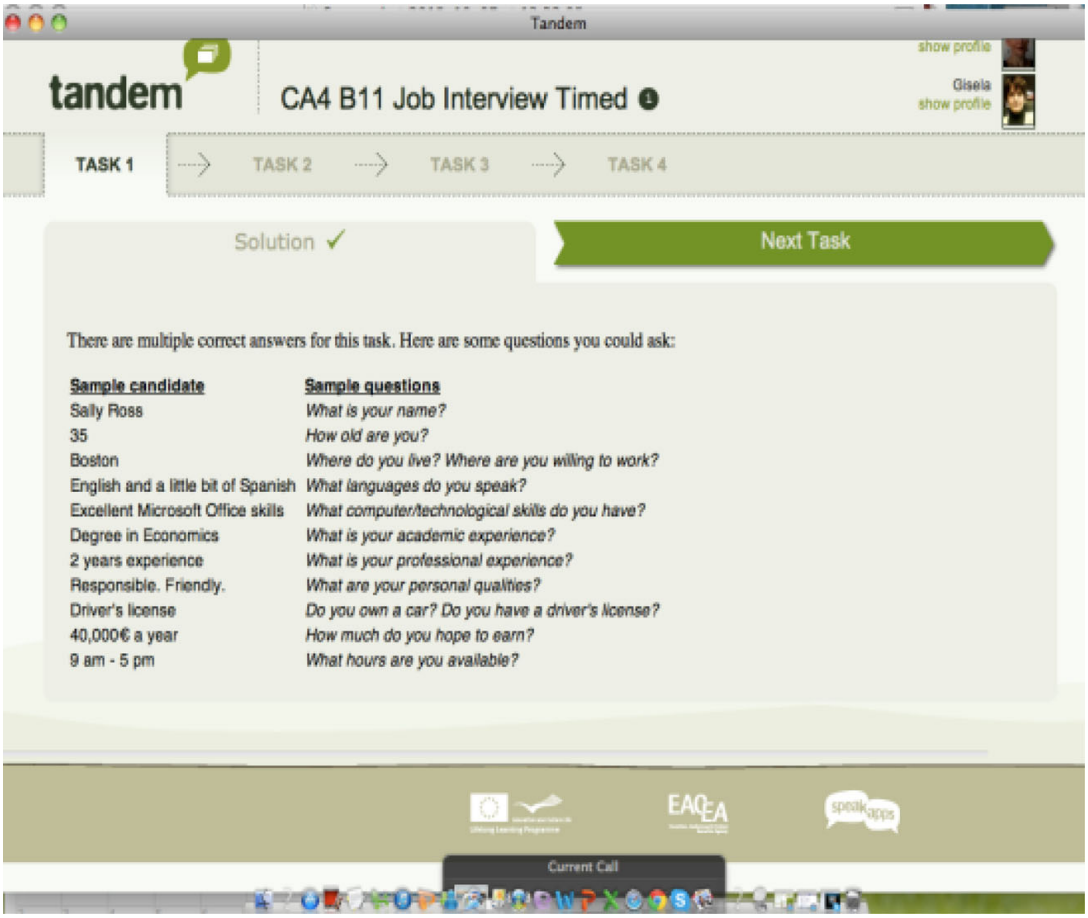
5.


tandem
CA4 B11 Job Interview Timed
show profile
TASK 1
TASK 2
TASK 3
TASK 4
See Solution
TIEMPO RESTANTE DE LA TAREA
01:16
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 ???
Current Call

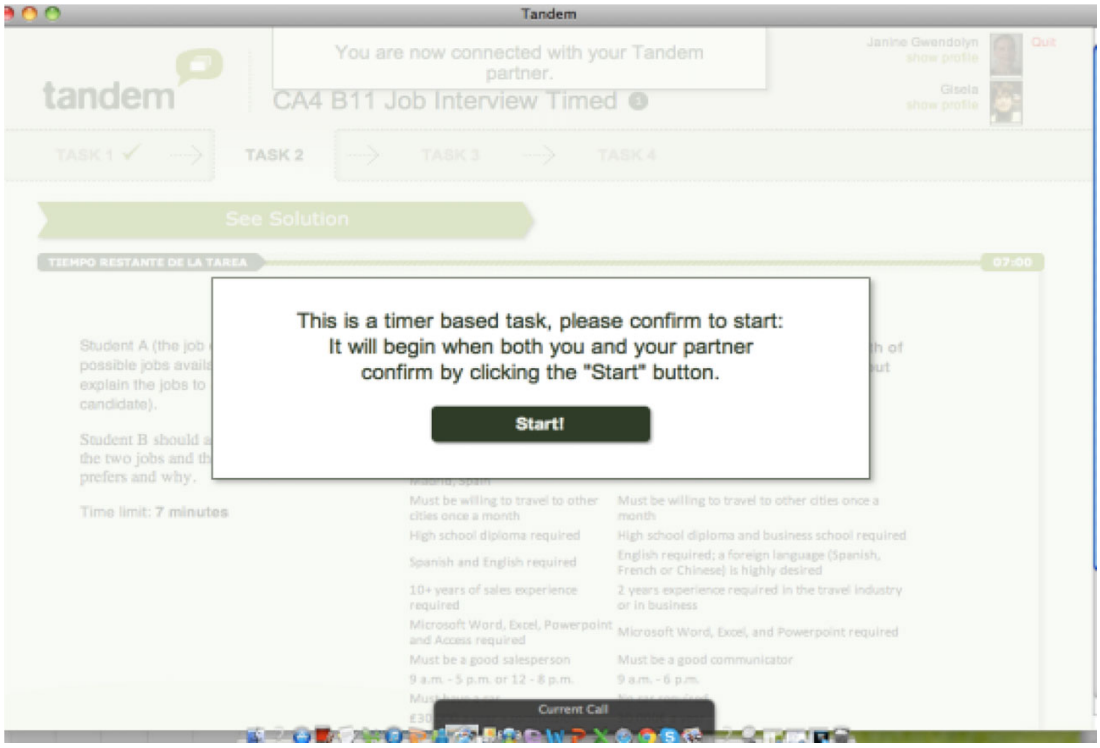
6.


tandem
CA4 B11 Job Interview Timed
TASK 1
TASK 2
TASK 3
TASK 4
Solution ✓
Next Task
There are multiple correct answers for this task. Here are some questions you could ask:
Sample candidate
Sally Ross
35
Boston
Time Up!
Close
Degree in Economics
2 years experience
Responsible. Friendly.
Driver's license
40,000€ a year
9 am - 5 pm
What are your personal qualities?
Do you own a car? Do you have a driver's license?
How much do you hope to earn?
What hours are you available?
Current Call

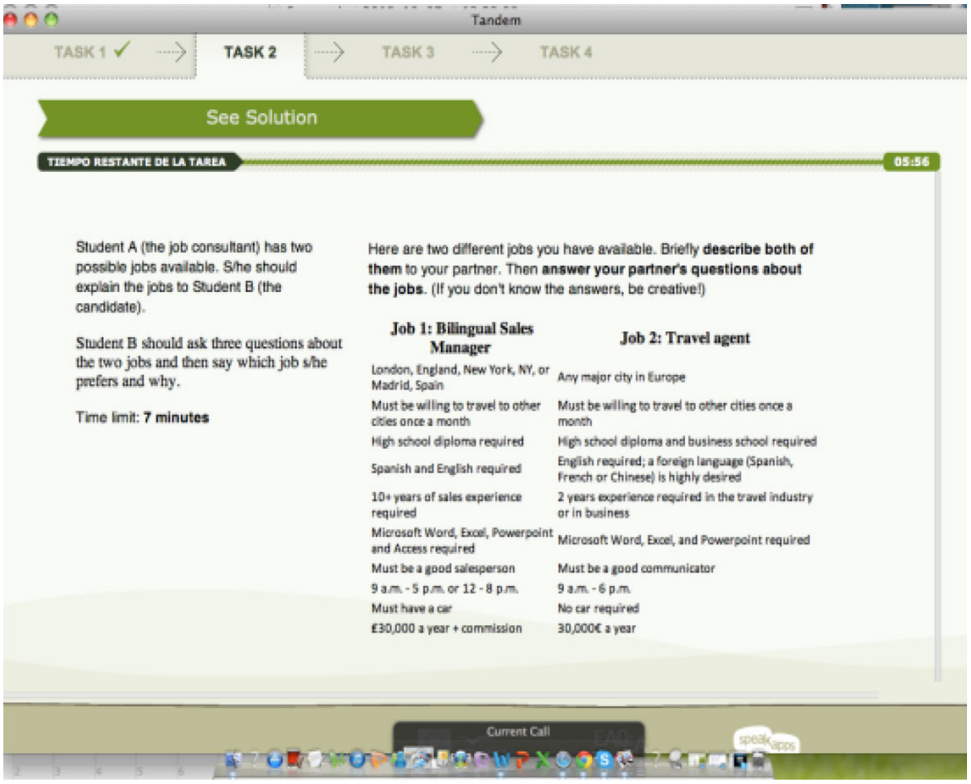
7.


Tandem
tandem
CA4 B11 Job Interview Timed
show profile
Gisela
show profile
TASK 1
TASK 2
TASK 3
TASK 4
Solution ✓
Next Task
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?
Current Call

8.


Tandem
You are now connected with your Tandem partner.
tandem
CA4 B11 Job Interview Timed
TASK 1
TASK 2
TASK 3
TASK 4
See Solution
TIEMPO RESTANTE DE LA TAREA
07:00
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.
Start!
Current Call

9.


Tandem
TASK 1
TASK 2
TASK 3
TASK 4
See Solution
TIEMPO RESTANTE DE LA TAREA
05:56
Student A (the job consultant) has two possible jobs available. S/he should explain the jobs to Student B (the candidate).
Student B should ask three questions about the two jobs and then say which job s/he prefers and why.
Time limit: 7 minutes
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
Current Call