

Managing multimodal data in virtual world research for language learning

CRISTINA PALOMEQUE

*University of Barcelona, Spain
(email: cristina.palomeque@gmail.com)*

JOAN-TOMÀS PUJOLÀ

*University of Barcelona, Spain
(email: jtpujola@ub.edu)*

Abstract

The study of multimodality in communication has attracted the attention of researchers studying online multimodal environments such as virtual worlds. Specifically, 3D virtual worlds have especially attracted the interest of educators and academics due to the multiplicity of verbal channels, which are often comprised of text and voice channels, as well as their 3D graphical interface, allowing for the study of non-verbal modes. This study offers a multilayered transcription method called the Multi-Modal MUVE Method or 3M Method (Palomeque, 2016; Pujolà & Palomeque, 2010) to account for the different modes present in the 3D virtual world of Second Life. This method works at two levels: the macro and the micro level. The macro level is a bird's-eye view representation of the whole session as it fits into one page. This enables the researcher to grasp the essence of the class and to identify interesting sequences for analysis. The micro level consists of three transcripts to account for the different communication modes as well as the interface activity that occurs in the virtual world of Second Life. This paper will review the challenges when dealing with multimodal analysis in virtual worlds and how the multimodal data were analyzed and interpreted by using a multilayered multimodal method of analysis (3M transcription). Examples will be provided in the study to show how different modes of communication were used by participants in the virtual world of Second Life to create meaning or to avoid communication breakdowns.

Keywords: multimodality, multimodal transcription, virtual worlds, Second Life

1 Introduction

Although communication has always been inherently multimodal, as there are several modes involved in any communicative event, linguistic research has often dealt with communication in a monomodal or bimodal way, initially focusing on the written mode and later incorporating spoken discourse (Kress & van Leeuwen, 2001). However, there are also other modes involved in communication that go beyond the verbal modalities. When these visual and spatial modes are not taken into account, it is difficult to fully understand the

communicative event (Kress, 1998; Kress & van Leeuwen, 2001). Today, these visual modes are gaining in relevance and one of the challenges in communication studies is how to account for and analyze the multiplicity of modes present in any communicative event and understand what role they play.

Multimodal data entail a different way of approaching the process of analysis, as there are several layers of information that should be taken into account. In order to tackle the complexity of dealing with this multiplicity of layers, this paper suggests a two-staged process. First, these layers should be analyzed individually, and in a second stage they will be analyzed as a whole in order to observe possible interactions that take place in the meaning-making process (Bezemer & Jewitt, 2010; Norris, 2004).

3D virtual worlds pose an additional challenge to the study of multimodal communication. In virtual worlds, there are further dimensions to take into account, such as avatar proxemics and spatial layout of the learning environment. Furthermore, participants manage their avatar and their communication needs through another layer, the interface.

This paper is based on a study that led to a PhD dissertation (Palomeque, 2016), which describes teacher and student communication in the 3D virtual world of Second Life. This study focuses on how the data management was carried out in order to analyze the multimodal interaction that occurred in a foreign language teaching-learning context.

This paper starts with a brief review of the approach that has grounded this study, including the relevance of analyzing multimodal discourse and using multimodal transcripts. It is followed by a description of the objectives and the context of the study. The next section presents an analysis of some of the challenges the researcher may encounter when dealing with multimodal data, ending with concluding remarks and future directions.

2 Multimodality and computer-mediated communication

The widespread use of social networking tools has offered promising affordances for computer-mediated communication (CMC) research, as the Internet is no longer solely a source for accessing information but also offers its users tools for communicating with other people around the world. Thus CMC is considered a popular area of research in computer-assisted language learning, as online communication tools are readily available and easy to access (Hubbard, 2009). Traditionally, many CMC studies (Lee, 2002; Pellettieri, 2000; Warschauer, 1995) have been text-based, focusing on the use of text chat or emails. However, with the introduction of multimodal tools, such as learning management systems, blogs, or virtual worlds, interest in CMC has been renewed as researchers have shifted their interest towards the nature of multimodal communication in such environments (Hampel & Hauck, 2006; Lamy, 2006; Tan, O'Halloran & Wignell, 2016; Wigham, 2012).

This paper takes a social semiotic approach to communication and focuses on how meaning is made through the interaction of several modes (Bezemer & Jewitt, 2010). Multimodal CMC is interpreted as the use of different modes of communication available to the user in an online environment (Herring, 2002). There are different types of multimodal CMC depending on the prevalent mode of communication available in the online environment. Thus multimodal CMC can be based on text, audio, video, or graphics (Herring, 2015). Although Second Life finds itself in the graphic-based category, there are several modes operating in this virtual world as it also contains text, audio, and video features.

Hence, the array of communication tools present in 3D virtual worlds such as Second Life offers promising affordances as they allow for a richer interactional experience where learners can select and use a variety of tools to suit their communicative purposes. Nevertheless, some researchers have also cautioned that this diverse availability of tools can also overwhelm learners and produce communicative breakdowns (Cunningham, Beers Fägersten & Holmsten, 2010; Tan *et al.*, 2016).

2.1 *Virtual worlds*

Virtual worlds have evolved enormously in the past years, starting from text-based MOOs, such as schMOOze University (Downey, 2014; Healey, 2016), to 3D virtual worlds, such as Second Life, which combine images and text and have incorporated a voice element as well as a stronger sense of presence due to more complex graphics and avatars.

Virtual worlds have sparked the interest of academics who have researched different aspects of language learning in these embodied environments. Wigham (2012), for instance, studied the verbal and non-verbal behavior of students in foreign language classes in Second Life and how students and teachers created meaning by using different modes. Peterson (2008) studied the discursive strategies that students used in a task-based setting in a virtual world and whether strategy use was influenced by task. Wang (2015) explored student participation in Second Life through a task-based approach.

One of the defining and most characteristic features of virtual worlds is the spatial dimension (Book, 2004) and the sense of presence the user has through their avatars (Friedman, Steed & Slater, 2007). Salmon (2009) points out that virtual worlds are ideal spaces for experiential and immersive learning due to their graphic interface. These features make 3D virtual worlds an attractive environment for online interaction. Social norms such as interpersonal distance have been found to be transferred from face-to-face contexts to virtual environments (Friedman *et al.*, 2007; Yee, Bailenson, Urbanek, Chang & Merget, 2007). Thus the 3D synchronous voice feature, as well as its strong sense of presence, makes virtual worlds a promising arena for online learning (Sweeney, Palomeque, González, Speck, Canfield, Guerrero & MacKichan, 2011).

Several authors have researched different features that contribute to the feeling of presence found in virtual worlds like Second Life, such as the study of interpersonal distance (Yee *et al.*, 2007) or the use of deictics or spatial cues to make reference to the virtual world (Örnberg, 2005; Wigham, 2012). Warburton (2009) describes Second Life as having three layers: the status, the communication, and the physical layer. The physical layer offers a visual element that consists of the physical environment and the avatars, who can adjust their proxemic behavior in the environment. The communication layer refers to the fact that participants can use voice or text communication tools, which are distance relevant. These layers contribute to the feeling of presence that the user experiences when logged into the virtual world (Warburton, 2009).

2.2 *Multimodal discourse analysis*

The main research instrument used in this study was computer-mediated discourse analysis. Although all discourse is multimodal, discourse analysis has tended to focus on verbal

communication, often overlooking non-verbal communication. However, this approach offers only a partial view of the communicative event (Norris, 2004). Multimodal discourse analysis (MDA) may help provide a more balanced view of communication.

The main tenet behind MDA is that online communication has its own features that are different from face-to-face communication (Herring, 2004). Thus medium variables such as channels of communication and their synchronicity will have an impact on the language that is generated (Herring, 2007).

In MDA, any mode has the potential of being more relevant than another in a given communicative event. Hence, MDA studies the modal configuration of a communicative event and how participants make use of the different available modes (Kress & van Leeuwen, 2001; O'Halloran, 2011). One of MDA's main concerns is *intersemiosis*, which explores the intermodal relations that result from the interaction of the participants' semiotic choices (Norris, 2004; O'Halloran, 2011).

2.2.1 Multimodal transcripts as tools for MDA. Traditionally, CMC studies have transcribed and analyzed interactions in a similar way as in a face-to-face class. However, when analyzing virtual worlds, these transcriptions become insufficient because, although many features are shared, some features are irrelevant and others are intrinsic to virtual worlds such as avatar movement or teleporting to different locations (Palomeque, 2016). Hence, there is a need to develop a multimodal transcript, a transcript that takes into account modes other than the verbal in the communicative event.

Several authors have provided examples of multimodal transcripts, focusing on different modes or multimodal texts. O'Halloran (2004), for example, carries out a multimodal analysis of films, Williamson (2007) studies the multimodality in written press, and Bearne (2009) has focused on the multimodality in storytelling. Authors offering multimodal transcripts have included different modes in their transcriptions depending on the purpose of their study. Baldry (2000), for example, analyzes a car advertisement and suggests a multimodal transcription that shows the interplay of different resources such as speech, sound, gesture, and writing. Domingo (2011), when analyzing digital video, includes a wide range of modes such as landscape, gestures, written and spoken language, visual and sound effects, and color. Tan *et al.* (2016) offer a notation system called Multimodal Analysis Video to analyze communication in face-to-face contexts as well as in Second Life. This software consists of a player window, a transcription box, and system strips for the semiotic modes of gaze, gesture, and virtual space. Thus multimodal transcripts can help analyze how meaning is created through the interaction of different modes and thus help overcome the limitations that face-to-face transcription conventions present.

Although some authors incorporate non-verbal information in their transcripts, Norris (2004) makes a case for including images from the communicative event in order to provide a richer description of the event and obtain nuances that may be overlooked in a verbal description. Norris (2004) points out that multimodal transcripts might entail several stages: in the first stage, the different modes are kept separate, and in a subsequent stage, the modes are viewed in a combined manner to analyze how meaning is created through interrelations. Multimodal transcriptions can be very labor intensive, and thus it may not be feasible or necessary to transcribe a whole lesson but to select relevant instances to analyze (Bezemer & Jewitt, 2010). Like Norris (2004), Bezemer and Jewitt's (2010) transcription

method consisted of a first phase where the modes were analyzed separately and of a second phase in which these modes were analyzed together.

3 Objectives

This study focuses on how the multimodal data gathered from the virtual world of Second Life was collected and analyzed. The objectives of the paper are to

- describe how the multimodal data were managed in terms of collection and analysis,
- highlight the challenges encountered when dealing with multimodal data,
- exemplify the different interconnected layers of multimodal data.

Before focusing on the core issue of this paper and dealing with how the multimodal data were managed and analyzed, it is relevant to clarify the objectives that this study is based on as well as provide information on the research context and how the data were collected.

4 Context of the research study

This paper is based on a research study (Palomeque, 2016) that focuses on the online discourse strategies, interactional modifications, and corrective feedback that the teacher uses in English for specific purposes lessons in Second Life from a multimodal perspective. The main objectives of this study were the following:

- To describe the teacher-generated discourse in a multi-user virtual environment (MUVE) context.
- To describe and analyze the teacher's use of online discourse strategies to manage a class in a MUVE context.
- To describe and analyze the communication modes used by participants and their functions in communication.
- To describe and analyze how the teacher manages a class in a MUVE by using different communication modes to achieve effective communication.

The research was carried out at the Escola Universitària d'Hoteleria i Turisme, a college that is attached to the University of Barcelona. The participants were undergraduate students of the degree in tourism and were enrolled in the first level of the subject English for Tourism Purposes. The initiation course corresponds to a B1 level, as per the Common European Framework of Reference for Languages. The main focus of these courses is for students to be able to cope with different situations linked to the tourism industry in English, and therefore the course content is highly functional and focused on oral fluency.

One group of the first level of English for Tourism Purposes was offered the opportunity to carry out additional fluency practice classes in the virtual world of Second Life in their spare time. Thirteen students (11 female and two male) volunteered to participate in the study. However, only nine finished the sessions as four had to abandon them because of technical reasons or work commitments.

These additional classes were composed of three modules, the content of which was linked to the general course syllabus. Each module contained three one-hour sessions in Second Life. The methodological approach in the design of the virtual sessions was task

Table 1. *Overview of the modules in the Second Life project (Palomeque, 2016)*

Module	Topic	Final task
Module 1	Hotels	Role play: Dealing with complaints about a hotel room
Module 2	Museums	Delivering a guided tour of a museum
Module 3	Cities	Delivering a guided tour around three important landmarks of a city

based, as the activities were geared towards the accomplishment of a final task at the end of each module. Furthermore, the purpose of the final tasks of each module was to deal with specific real-life situations that students would be likely to encounter in their future careers (Nunan, 1989). The final tasks had both a social/communicative as well as a spatial dimension (Deutschmann & Panichi, 2009), as the MUVE environment and the selected location played an essential role in the design and performance of the tasks. Table 1 illustrates the content and final tasks that the students performed.

Now that the context has been established, we will turn to how the multimodal data were collected.

5 Managing multimodal data collection

The main source of data for the study was the screen recordings of the sessions that took place in Second Life. However, students were also asked to fill in pre-, “during”, and post questionnaires to add significance to the recorded data.

First, a plot of land was rented and used as the headquarters for the class to meet at the beginning of each session. The participants were divided into two groups that met at different times, as having all of them in one session caused technical sound problems. Unfortunately, some of the recordings were lost due to technical factors. The first obstacle was that the Second Life program needs certain technical requirements to run smoothly such as a good graphics card and a fast Internet connection, which not all computers had. Furthermore, the screen recorder that was used, SnapzProX by Ambrosia Software Inc., produces extremely heavy high-quality .mov files, which are difficult to manage. The most delicate moment of the recording process happened at the end of the session when the program converted the video file, as sometimes the computer crashed and the session was lost. Other times, the screen recorder recorded only one of the channels, the microphone channel or the background sound. A total of 17 hours was collected. However, due to the extent of the recordings, the data object of analysis was reduced to one group, amounting to nine hours and thirty-five minutes.

Thus there is a wide range of technical problems that the researcher might have to deal with when collecting multimodal and multichannel data. The planning stage previous to the recording of the online encounters is essential to minimize problems and ensure the data collection.

5.1 The 3M method

In section 2.2.1, different multimodal transcriptions for different multimodal texts have been reviewed. In this section, we propose a multimodal transcription method that takes into

account some of the recommendations from the abovementioned section (Bezemer & Jewitt, 2010; Norris, 2004) in order to analyze the communication that takes place in the virtual world of Second Life.

This study offers a multilayered transcription method called the Multi-Modal MUVE Method or 3M Method (Palomeque, 2016; Pujolà & Palomeque, 2010) to account for the different modes present in the 3D virtual world of Second Life. The 3M Method operates at two levels: the macro level, where the scenes of a class are observed, and the micro level, in which the selected scenes are transcribed in a multilayered manner.

5.1.1 Macro level. At the macro level, contextual information is collected such as the course description, the list of participants, and location. The location, unlike in traditional face-to-face classes, is relevant in a class that happens in a virtual world, as it can take place in any setting that the teacher has selected or designed for the lesson. The macro level presents a bird’s-eye view of the whole class and provides the researcher with information about the development of the class. The class is portrayed as a table and it is read as a musical score, from left to right (see Figure 1).

The class is divided into scenes. Each new scene is marked by a significant camera shift or a location change. The scenes are numbered and their duration is specified in the row above. Transitions are marked by a change in scenario or continuous camera shift. Transitions last a few seconds and usually occur when participants are walking or teleporting from one location to another. The scenes have been grouped into classroom activities, which have a describing caption in the row above to help in the understanding of the class sequence and to help the researcher select the appropriate scenes. The session portrayed in Figure 1, for example, comprises four stages: warm-up, a hotel race, a check-in role-play activity, and leave takings.

One of the concerns when transcribing multimodal events involves tackling a large quantity of data. The researcher must make a series of decisions, such as selecting what fragments of the data will be transcribed, as often an in-detail analysis of all the data collected is too vast for the researcher (Bezemer & Jewitt, 2010). The macro layer provides the researcher with a holistic view of the lesson so that the relevant scenes can be selected according to the researcher’s purposes. For instance, the researcher may want to analyze


0:00:00 - 0:26:03				0:26:03 - 0:29:00											
Warm-up/SL skills				Activity 1: Hotel race											
0:00:00 – 0:10:02	0:10:0 2 –	0:10:1 3 –	0:26:0 3 –	0:26:3 6 –	0:26:5 8 –	0:27:0 4 –	0:27:2 1 –	0:27:2 8 –	0:27:4 2 –	0:27:5 4 –	0:27:5 9 –	0:27:5 3 –	0:28:2 8 –	0:28:4 8 –	0:29:0 0
S01	S02	S03	T01	S04	T02	S05	T03	S06	T04	S07	S08	T05	T06		
															
0:29:00 - 0:57:16													0:57:16 - 1:01:13		
Activity 2: Check-in role play													Leave takings		
0:29:00 –	0:34:26 –	0:34:36 –	0:44:50 –	0:49:37 – 0:51:16	0:51:16 – 0:52:20	0:52:20 – 0:55:01	0:55:01 – 0:57:16	0:57:16 – 1:01:13							
S09	S10	S11	S12	S13	S14	S15	S16	S17							

Figure 1. Example of the macro layer of a class session

how instructions are delivered at the beginning of each activity or he or she may be looking for instances of group work in order to analyze peer interaction. In Figure 1, the researcher is interested in the greetings and warm-up stage; therefore, a screenshot of the class is placed in the first cell corresponding to Scene 1.

The macro transcription has two aims: the first aim is to get the big picture of the lessons in relation to the teaching sequence, and the second one is to help the researcher make a selection of the most relevant scenes for the purpose of the study and which will be transcribed at the micro level.

5.1.2 Micro level. Once the researcher has selected the scenes from the macro level, she or he turns to the micro level. The micro level comprises two stages. The first stage consists of a scene description that includes a description of the location, the nature of the activity, and the participants who are present.

The second stage involves a multilayered transcription. The 3M transcription framework consists of a video player, which displays the class recording. Below the video player, there are three transcripts: two of them account for the communication that takes place through the different modes, and the third one deals with the interface activity (see Figure 2). The first transcription is devoted to the verbal mode. The verbal mode contains the oral and the written channel including public and private oral and written chat. In Figure 2, the first sender, the teacher (T), starts with an oral turn, and the following three turns are carried out by a student (SyHe) and the teacher through the local chat (LC). The second transcript is for the visual mode, where the transcriber includes avatar-generated actions such as gestures as well as environment-generated actions such as sitting or teleporting. In Figure 2, students carry out the environment-related action of standing around the fire in (00:28), and there are several avatar-related behaviors such as sitting (00:43) or jumping (01:23). The last transcription describes the participant's use of the

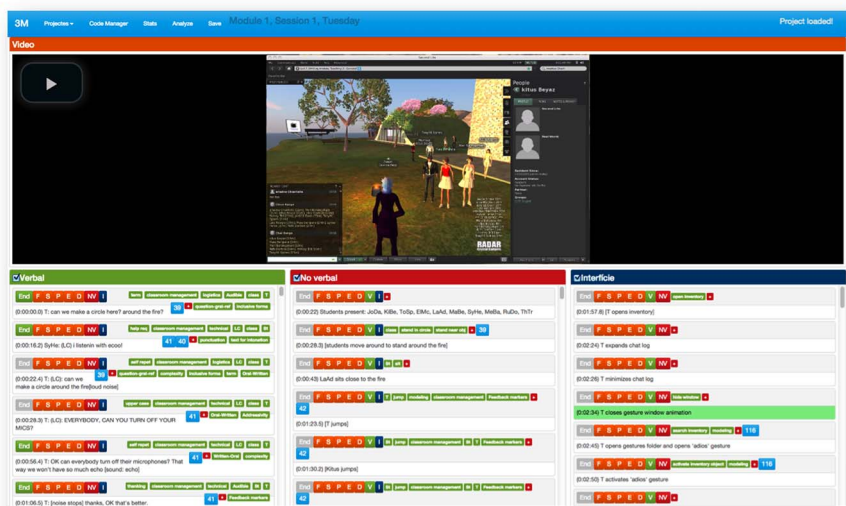


Figure 2. 3M software showing a micro transcription fragment of a class session

interface to manage Second Life such as window and inventory management. All the transcriptions are synchronized with the video of the class, which allows the researcher to visualize the video with the transcriptions in real time as well as to perform queries on the transcriptions to identify relevant sequences.

After transcribing the different modes, each turn and action is coded. When the turns or actions belong to the same discursive sequence, they are linked to each other by assigning the sequence with a number. Therefore, when a strategy is queried, the researcher will see the whole discursive sequence, thus obtaining the turns that are involved in the sequence regardless of the transcripts. This enables the researcher to observe the turns in a more contextualized and holistic way and view all the modes that are involved in the meaning-making at the same time. In Figure 2, it can be seen that sequence 39 has verbal and visual turns. The teacher uses both the oral and the text channels in the verbal model to ask students to get into a circle around the fire, and the students respond visually by gathering around the fire.

6 Managing multimodal data analysis

The researcher must tackle a number of challenges when dealing with multimodal data, including issues such as quantity, variety of channels, and the multilayered nature of a multimodal transcription; what happens in-world is triggered by actions that the users perform on the interface. This section will review some of the challenges when analyzing multimodal data. The first part deals with deciding on the unit of analysis. The second challenge tackles the multilayered nature of the virtual world, including accounting for the in-world as well as the interface dimension and studying the meaning-making process through the codependence of different modes that intervene in-world.

6.1 Unit of analysis

The unit of analysis is an essential decision a researcher must make when transcribing interaction, and it will depend on the focus of the research. In her study, Wigham (2012) analyzes learners' and teachers' verbal and non-verbal behavior in Second Life and uses "acts" as her unit of analysis. Units of analysis become more complex when dealing with multimodal data because of the multiplicity of layers, modes, and channels that are involved. Baldry and Thibault (2001) use instance and type in their transcriptions, and Baldry (2004) also explores the relationship of phases and transitions in multimodal texts.

The present study has used several units of analysis at different levels. At a macro level, the units of analysis operate as in a face-to-face teaching-learning context. The first unit of analysis is the teaching sequence. Within a teaching sequence the teacher sets into motion a series of learning activities with the aim of generating interaction (Cambra Giné, 2003). The teaching sequence is made up of a series of sessions, the second unit of analysis, which has its aims, contents, and learning activities.

One of the aims of the research study (Palomeque, 2016) was to analyze discursive strategies used by learners and the teacher. Thus the units that were used were the turn for verbal interactions, where *turn* is understood as "each time the 'floor' [is] transferred from one participant to another, regardless of its length" (Tudini, 2003: 148), and the action for non-verbal behavior. The term "action" is taken from Norris's (2004) concept of

“embodied mode” where a participant uses a series of different modes of communication through the orchestration of higher level and lower level actions. The last unit of analysis was the discursive sequence, which is made up of all the turns that intervened in a communication strategy such as the trigger of a communication strategy and the reaction (Palomeque, 2016). In Figure 2, sequence 41 consists of a student, SyHe, who is complaining because of sound problems. This triggers a reaction from the teacher, and she asks the students through the oral and the written channels to turn off their microphones.

6.2 Multilayered communication

Transcribing different modes and different layers of a communicative event can help us obtain a richer analysis and understanding of what is occurring. However, including all the modes, channels, and layers in one transcription can make the final analysis confusing. One possible way to unravel this complexity is to analyze the data in two stages (Bezemer & Jewitt, 2010). In the first stage, the researcher will transcribe the different modes individually and analyze them separately. Also, a separate transcript will be allocated to interface actions; that is, the actions that the user performs on the program that cannot be seen in-world. In a second stage, through an interdependent tagging system, the researcher can tag and link turns and actions that belong to the same sequence but appear in different transcripts in order to analyze the data in a more holistic and connected way.

The following sections will review different types of modal interdependence that the researcher may encounter.

6.2.1 In-world interface interdependence. Participants can navigate the world thanks to an interface that allows users to manage their avatar and perform actions such as using the microphone, text chat, teleporting to another location, or activating an object in-world. Transcribing the teacher’s interface actions, for instance, can help understand the technical skills as well as classroom management strategies that a teacher needs to deploy in a virtual world.

The following example illustrates three different transcripts together. Due to the constraints of traditional delivery on paper, the turns and actions appear here in sequential mode. The different modes are specified in brackets and indented. Some modes are highlighted in the examples to help observe the modal interdependence. For instance, the interface actions have been highlighted in bold in Excerpt 1. This complexity is overcome with multimodal online software such as the 3M program (see Figure 2), which facilitates the transcribing process, the reading of the data, and its subsequent analysis.

Excerpt 1. *Example of in-world interface interdependence*

(interface) (0:03:33) T types “hey!” to activate waving gesture

(verbal) (0:03:33) T: can you try some other gestures?

(visual) (0:03:33) T makes a waving gesture

(interface) (0:03:38) T opens inventory and activates clapping gesture

(visual) (0:03:38) T makes clapping gesture

(interface) (0:03:48) T types “hey!” to activate waving gesture

(visual) (0:03:48) T makes a waving gesture

(visual) (0:03:54) JoDa makes gesture: “get lost!”

(visual) (0:03:55) MeBa makes gesture: “get lost!”

(verbal) (0:03:57) T laughs

Excerpt 1 illustrates an example of the first session of the virtual tourism project. The teacher has shown the students how to activate the gestures in their inventory and is asking them to practice. We can notice how there are several actions from the three different transcripts (the interface, the verbal, and the visual) occurring at the same time: while the teacher is asking the students to try other gestures, she activates a gesture in her inventory, and, as a result, her avatar is animated and waves. The same happens with the clapping gesture and again with the waving gesture. This synchronicity of linguistic and other semiotic systems to create meaning has been reported in other computer-assisted multimodal settings, such as learners using the Lyceum multimodal environment (Lamy, 2006). Finally, two students try a new gesture; JoDa tries the gesture “get lost”, and MeBa does the same. As a reaction to what has occurred in the visual mode, the teacher laughs at their gestures. This sequence could not be understood without the presence of the three different transcripts.

6.2.2 Modal interdependence within the virtual world. One of the main concerns in MDA is *intersemiosis*. This section includes different examples of how different modes or channels are used in combination to create meaning in-world. The first section involves modal interdependence within the verbal domain, which includes the audible and the textual modalities. The second section studies the interdependence of the verbal and the visual mode.

(a) Verbal interdependence (audio and text channels)

As seen in section 5.1, in the 3M transcription, one of the transcripts is devoted to the verbal mode. However, this transcript encompasses two channels: the text and the audio. Furthermore, within each channel there are different types of communication. For instance, within the text channel, students can send private instant messages or use the public local chat. Thus one transcript contains several channels of communication as well as different subtypes depending on the number of participants involved in the interaction or depending on whether the channel is public or private. The different channels and subtypes were specified in brackets after the sender of the message and also coded separately to facilitate future queries.

When observing the verbal transcript, many instances were found where participants made use of both the audio and the text channels in combination to communicate more effectively. Students made use of channel switches for several reasons. In particular, channel switches occurred as a technical compensation strategy, such as when someone was encountering problems with their audio or microphone.

Excerpt 2. *Example of verbal interdependence of the audio and text channels*
(Palomeque, 2016)

(verbal – audio) (0:30:35) JoDa: San Francisco is ???

(verbal – audio) (0:30:37) T: Sorry? San Francisco is?

(verbal – audio) (0:30:41) JoDa: eh: ???

(verbal – audio) (0:30:44) T: sorry, JoDa, I can't I can't understand you. Uhm is it over?

(verbal – local chat) (0:30:47) T (LC): is it over?

(verbal – local chat) (0:30:49) JoDa (LC): that's all

In the example shown in Excerpt 2, the student JoDa is giving a guided tour of virtual San Francisco. At this point, he starts having microphone problems and cannot be understood. The teacher repeats her question through the text channel to make sure that JoDa gets her message. JoDa also switches to the text channel and repeats what he said using the local text chat to clarify to the rest of the participants what he was trying to say through the oral channel.

Furthermore, channel switches were also used for other purposes; for instance, the local chat was also used as a means to gain a participant's attention.

Excerpt 3. *Example of verbal interdependence of the audio and text channels*
(Palomeque, 2016)

(verbal – audio) (0:39:56) RuDo: Ma, we: look around and and looking for information or ...?

(verbal – local chat) (0:40:16) RuDo (LC): maiaaa???

(verbal – audio) (0:40:20) MaBe: yes

In the example that appears in Excerpt 3, students RuDo and MaBe are working together to prepare their guided tour. RuDo asks MaBe a question over the voice chat but gets no reply after waiting for 20 seconds. Although RuDo's microphone is working, she tries another channel to call her partner, to which MaBe replies.

The teacher often used channel switches to deal with technical problems, as did the students, but she also used them for teaching purposes such as to summarize class instructions, to model new words, and to show presence in an unobtrusive way when students were holding the floor, as can be seen in the example shown in Excerpt 4.

Excerpt 4. *Example of channel switch to show presence unobtrusively* (Palomeque, 2016)

(verbal – audio) (0:29:56) ArCh: here we can see a very colorful floor. And there are not a lot of things ah: but the carpet is blue a: with the yellow submarine.

(verbal – audio) (0:30:10) T: mhm

(verbal – audio) (0:30:12) ArCh: there are also, there is also a fa: uhm a poster of the Beatles with the famous words of the song. All you need is love.

(verbal – local chat) (0:30:27) T (LC): great

There is evidence of verbal interplay in the data with different purposes. However, the main purposes of the channel switches, as can be seen in the examples, show the chat being

used as a means to complement or compensate for the audio channel (Cunningham *et al.*, 2010; Hampel & Stickler, 2012). These channel switches proved to be a useful communication strategy for both the students and the teacher to maximize communication in this multichannel environment and compensate for communication problems. Also, channel switches can be an effective classroom management strategy in these environments as a way of providing unobtrusive feedback or scaffolding to learners and of ensuring that the message reaches the students (Palomeque, 2016).

(b) Verbal and visual interdependence

As seen in the literature review, the sense of presence that 3D virtual worlds, such as Second Life, provide has an effect on the interaction that is produced in such worlds (Örnberg, 2005). The visual mode is a defining trait of 3D virtual worlds and, hence, must be taken into account in a multimodal analysis of communication.

Although several studies report a strong sense of presence (Friedman *et al.*, 2007; Yee *et al.*, 2007), there were few conscious avatar-related non-verbal behaviors found in the study. Gestures were seldom used by students, probably due to the fact that a user has to consciously activate a gesture in a virtual world, limiting the spontaneity of the moment. In face-to-face contexts, the verbal and non-verbal modes are often synchronized. However, in Second Life, students have two learning curves to overcome: using a new software program plus learning a foreign language. As a result, students may experience a cognitive overload (Anderson, 2009; Atkins & Caukill, 2009; Palomeque, 2016; Sweeney *et al.*, 2011).

Despite the little evidence found concerning gestures, the visual mode played an important role in communication as a means of avoiding ambiguity, especially in context-dependent tasks such as the guided tour or for non-verbal responses.

Excerpt 5. *Example of verbal and visual interdependence (Palomeque, 2016)*

(verbal – audio) (0:00:00) T: Can we make a circle here? Around the fire?

(verbal – local chat) (0:00:16) SyHe (LC): I listen with ecoo!

(verbal – local chat) (0:00:22) T (LC): can we make a circle around the fire?

(visual) (0:00:28) Students move to stand around the fire

In Excerpt 5, participants are logging into Second Life to start the class, and they are walking around the location. The teacher asks the students through the oral channel to make a circle around the fire. However, some students are having technical problems, so she switches to the text channel and repeats her request. The students respond non-verbally by forming a circle around the fire.

This interdependence of the verbal and the visual mode can be seen by the use of strategies such as performing location checks, referring to in-world objects, or adjusting the position of the user's avatar when interacting with another avatar. The following examples are the results of queries that combine the verbal and the visual mode. Although they belong to different transcripts, they have been tagged as pertaining to the same sequence.

The visual mode was very important to ensure that all the students were at the same location. Participants often used location checks whenever there was a location change

during the lesson, as sometimes not all students arrived at the target destination either because of technical problems caused by their lack of familiarity with the program or because of the instability of the program.

Excerpt 6. *Example of verbal and visual interdependence (Palomeque, 2016)*

(interface) (0:41:13) T opens San Francisco landmark guide

(verbal – audio) (0:41:16) T: OK, girls, see you there. You can uhm teleport us when you get there.

(interface) (0:41:44) T teleports to Golden Gate

(verbal – audio) (0:43:29) T: Wait, let's teleport Me. Where's MeBa?

(visual) (0:43:33) T walks around the area and stops in front of NoLe and KeHu

(interface) (0:43:38) T looks for MeBa in her friend list and sends her a tp offer

(visual) (0:44:00) MeBa teleports to Golden Gate and joins group

Excerpt 6 shows an extract where the class teleports to a new location: the virtual Golden Gate. After performing a visual check, the teacher realizes that a student, MeBa, is missing so she sends her a teleport offer. Location checks were an important classroom management strategy and were used whenever there was a location change to make sure that all the participants had teleported to the new location successfully. This strategy was also used by students when they were in charge of their guided tour.

The teacher and students often needed to refer to the shared virtual space during the lesson. There were many verbal in-world references with a logistical or technical purpose, used to clarify certain points of the lesson.

Excerpt 7. *Example of verbal in-world references (Palomeque, 2016)*

(verbal – audio) (0:57:51) T: Ne ... where's NoLe? Did she fly away?

(visual) (0:57:53) MeBa and KeHu stop flying and walk to the teacher's location

(verbal – audio) (0:58:04) T: Where's NoLe? Wait, can somebody teleport her? I think she ... let's come here for a second?

(verbal – audio) (0:58:12) KeHu: NoLe:!

(verbal – audio) (0:58:16) T: I see NoLe flying

(verbal – audio) (0:58:19) T: NoLe!

(verbal – audio) (0:58:22) NoLe: I'm lost!

(verbal – audio) (0:58:23) T: here, here, just uhm we can see you from here. We're in the first house.

(visual) (0:58:35) NoLe lands and joins the group

In Excerpt 7, the group is flying and exploring the houses in Lombard Street, but NoLe flies away. The teacher, using the deictic “here”, tells the group to come to where she is standing. MeBa and KeHu come, but NoLe keeps flying. The teacher then repeats her request using the same deictic and using a term to specify their location. In this example, we can see how the visual mode and the verbal mode are used in combination to clarify the location, as “here” without the visual of the teacher’s avatar’s location would be meaningless.

Regarding non-verbal in-world references, Wigham (2012) found that participants used deictic gestures and avatar movement as a means of contextualizing the deictics in the environment. In this study (Palomeque, 2016), participants used very few deictic gestures such as pointing, and avatar proximity was frequently used as a strategy to make reference to in-world objects.

Excerpt 8. Example of proxemic in-world references (Palomeque, 2016)

(verbal – local chat) (1:06:30) T (LC): do you know what the instrument in the corner is?

(verbal – local chat) (1:06:36) T (LC): this long guitar?

(visual) (1:07:01) T walks to the sitar

In Excerpt 8, the students are in a room full of musical instruments and the teacher walks towards the instrument she is making reference to in order to avoid ambiguities.

The codependence of the visual and the verbal mode was also observed when avatars were interacting with each other. Avatar proxemics, facing another avatar when engaging in interaction, was found to be closely linked to verbal addressivity. Addressivity, expliciting the intended addressee of a message, is an important strategy identified in CMC research (Werry, 1996) to avoid ambiguity concerning the addressee of a message. However, few studies have researched addressivity in embodied environments such as 3D virtual worlds. Naper (2011) and Peterson (2008) found a low rate of explicit addressivity and point out that the possibility of avatars being able to move closer to each other may explain this. However, Wigham (2012) found that avatars did not orientate their avatars towards each other when communicating.

In this study (Palomeque, 2016), most instances of verbal addressivity were accompanied by a proxemic component; that is, avatars tended to face each other when engaging in interaction. Thus addressivity in Second Life was both verbal and visual. Sometimes the visual addressivity was conditioned by the learning space designed by the teacher. For example, at the beginning and end of the lesson, participants sat on a carpet that had animated cushions that made the avatars face each other. However, when exploring other locations or engaging in other activities, avatars tended to naturally join a circle or come closer to other avatars.

Excerpt 9. Example of visual addressivity (Palomeque, 2016).

(visual) (0:37:33) T teleports to Virtual Hallucinations Museum

(visual) (0:37:33) T has her back to students

(visual) (0:37:49) MeBa teleports to Virtual Hallucinations Museum

(verbal – audio) (0:37:55) T: Me, I think that:, wait, let’s see if NoLe can hear us, if not it’s only you. Wait, I teleported NoLe,

(verbal – audio) (0:38:33) T: Me, can you hear me?

(visual) (0:38:35) T faces students

(visual) (0:38:35) Students form a circle

In Excerpt 9, the teacher and Me have just teleported to a new location and the teacher starts talking to Me, although her avatar is not facing her. However, in her second turn, the teacher adjusts her avatar’s position to face Me and students naturally form a circle as they arrive at the new location. It was observed that it sometimes took longer to adjust the user’s avatar to face-to-face proxemic norms such as facing the interlocutor due to the fact that the in-world graphics take time to load and it takes time for the user to adjust to the new location (Palomeque, 2016).

These examples show how the visual mode played an important role in communication. Furthermore, apart from having a transcript for the visual mode, having the video of the lesson synchronized with the different turns and actions helped the researcher gain a more complete understanding of the different sequences.

7 Conclusion

The aim of the 3M Method is to provide an annotation system that integrates both the verbal and the visual modes to obtain a better understanding of communication in a multimodal virtual world. By having separate transcripts, the method allows the researcher to focus first on the different modes separately, and, after the coding process, the researcher can observe how the different modes work in combination to create meaning.

The use of the 3M Method in this study allowed the researcher to observe how participants created meaning through the use of different modes and channels. Furthermore, the combined use of different modes allowed participants to communicate more effectively and avoid ambiguities. This finding is in line with Lamy (2006) who states how learners used the different available channels and modalities to achieve their discursive aims. Also, we can see how the visual mode was important in Second Life through both the frequent, explicit reference to the visual mode by the participants and the proxemic organization of its participants. Thus this mode must be taken into account when transcribing interaction in a 3D virtual world. Not only is multimodality important for researchers but also students and teachers need to become familiar with the different modes and channels available, as well as their affordances, in order to use them effectively in a lesson in a virtual world. Hampel and Hauck (2006: 12) state that “language learners will have to become competent in both switching linguistic codes *and* switching semiotic modes and do so consciously. On top of that, they have to become ‘fluent’ in new codes such as online speech and writing *and* image”.

This study has its limitations, as the only screen recordings available are the teacher’s. In order to get a more holistic understanding of the interaction process, it is important to obtain the students’ screen recordings as well. This would enable the researcher to observe the actions participants perform on the interface as well as the private messages that are sent. Another dimension that could be relevant to study is the on-screen and off-screen behavior of the participants (Pujolà, 2002), as the screen recordings only show the on-screen actions.

Although nowadays 3D virtual worlds such as Second Life have lost their popularity, more studies are needed to analyze how learners and teachers manage the multiplicity of modes in different multimodal online environments. Several authors have pointed to the low participation rate of students in multimodal environments (Hampel & Stickler, 2012; Wang, 2015). Thus more studies are needed on teacher strategies to promote participation through the different channels or modes available in multimodal environments. These studies will certainly need a methodological approach to manage data in a multilayer way similar to the one presented in this paper.

References

- Anderson, T. L. (2009) Online instructor immediacy and instructor-student relationships in second life. In Wankel, C. & Kingsley, J. (eds.), *Higher education in virtual worlds: Teaching and learning in Second Life*. Warrington: Emerald Group Publishing, 101–114.
- Atkins, C. and Caukill, M. (2009) Serious fun and serious learning: The challenge of Second Life. In Molka-Danielsen, J. & Deutschmann, M. (eds.), *Learning and teaching in the virtual world of Second Life*. Trondheim: Tapir Academic Press, 79–89.
- Baldry, A. (2000) Introduction. In Baldry, A. (ed.), *Multimodality and multimediality in the distance learning age: Papers in English linguistics*. Campobasso: Palladino Editore, 11–39.
- Baldry, A. and Thibault, P. J. (2001) Towards multimodal corpora. In Aston, G. & Burnard, L. (eds.), *Corpora in the description and teaching of English*. Bologna: CLUEB, 87–102.
- Baldry, A. P. (2004) Phase and transition, type and instance: Patterns in media texts as seen through a multimodal concordancer. In O'Halloran, K. L. (ed.), *Multimodal discourse analysis: Systemic functional perspectives*. London: Continuum, 83–108.
- Bearne, E. (2009) Multimodality, literacy and texts: Developing a discourse. *Journal of Early Childhood Literacy*, **9**(2): 156–187. <https://doi.org/10.1177/1468798409105585>
- Bezemer, J. and Jewitt, C. (2010) Multimodal analysis: Key issues. In Litosseliti, L. (ed.), *Research methods in linguistics*. London: Continuum, 180–197.
- Book, B. (2004) Moving beyond the game: Social virtual worlds. *Proceedings of State of Play 2 Conference*. New York: New York Law School, 6–8.
- Cambra Giné, M. (2003) *Une approche ethnographique de la classe de langue*. Paris: Didier.
- Cunningham, U., Beers Fägersten, K. and Holmsten, E. (2010) “Can you hear me, Hanoi?” Compensatory mechanisms in synchronous net-based English language learning. *The International Review of Research in Open and Distance Learning*, **11**(1): 161–177. <https://doi.org/10.19173/irrodl.v11i1.774>
- Deutschmann, M. and Panichi, L. (2009) Instructional design, teacher practice and learner autonomy. In Molka-Danielsen, J. & Deutschmann, M. (eds.), *Learning and teaching in the virtual world of Second Life*. Trondheim: Tapir Academic Press, 27–44.
- Domingo, M. (2011) Analyzing layering in textual design: A multimodal approach for examining cultural, linguistic, and social migrations in digital video. *International Journal of Social Research Methodology*, **14**(3): 219–230. <https://doi.org/10.1080/13645579.2011.563619>
- Downey, S. (2014) History of the (virtual) worlds. *The Journal of Technology Studies*, **40**(2): 54–66. <https://doi.org/10.21061/jots.v40i2.a.1>
- Friedman, D., Steed, A. and Slater, M. (2007) Spatial social behavior in Second Life. In Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K. & Pelé, D. (eds.), *Intelligent virtual agents: 7th International Conference, IVA 2007, Paris, France, September 2007 Proceedings*. Berlin Heidelberg: Springer, 252–263. https://doi.org/10.1007/978-3-540-74997-4_23
- Hampel, R. and Hauck, M. (2006) Computer-mediated language learning: Making meaning in multimodal virtual learning spaces. *The JALT CALL Journal*, **2**(2): 3–18.

- Hampel, R. and Stickler, U. (2012) The use of videoconferencing to support multimodal interaction in an online language classroom. *ReCALL*, **24**(2): 116–137. doi: <https://doi.org/10.1017/S095834401200002X>.
- Healey, D. (2016) Language learning and technology: Past, present and future. In Farr, F. & Murray, L. (eds.), *The Routledge handbook of language learning and technology*. Abingdon: Routledge, 9–22.
- Herring, S. C. (2002) Computer-mediated communication on the internet. *Annual Review of Information Science and Technology*, **36**(1): 109–168. <https://doi.org/10.1002/aris.1440360104>
- Herring, S. C. (2004) Computer-mediated discourse analysis: An approach to researching online behavior. In Barab, S. A., Kling, R. & Gray, J. H. (eds.), *Designing for virtual communication in the service of learning*. Cambridge: Cambridge University Press, 338–376. <https://doi.org/10.1017/CBO9780511805080.016>
- Herring, S. C. (2007) A faceted classification scheme for computer-mediated discourse. *Language@Internet*, **4**: 1–37.
- Herring, S. C. (2015) New frontiers in interactive multimodal communication. In A. Georgopoulou & T. Spilloti (eds.), *The Routledge handbook of language and digital communication*. London: Routledge, 398–402.
- Hubbard, P. (2009) General introduction. In Hubbard, P. (ed.), *Computer assisted language learning: Critical concepts in linguistics*. London: Routledge, 1–20.
- Kress, G. (1998) Visual and verbal modes of representation in electronically mediated communication: the potentials of new forms of text. In I. Snyder & M. Joyce (Eds.), *Page To Screen: taking literacy into the electronic era* (pp. 53–79). London: Routledge.
- Kress, G. and van Leeuwen, T. (2001) *Multimodal discourse: The modes and media of contemporary communication*. London: Arnold.
- Lamy, M.-N. (2006) Multimodality in online language learning environment: Looking for a methodology. In Baldry, A. and Montagna, E. (eds.), *Third International Conference on Multimodality*, Pavia: Palladino, 237–254. <https://telearn.archives-ouvertes.fr/hal-00197411>
- Lee, L. (2002) Synchronous online exchanges: A study of modification devices on non-native discourse. *System*, **30**(3): 275–288. [https://doi.org/10.1016/S0346-251X\(02\)00015-5](https://doi.org/10.1016/S0346-251X(02)00015-5)
- Naper, I. (2011) Conversation in a multimodal 3D virtual environment. *Language@Internet*, **8**(7). <http://urn:nbn:de:0009-7-32225>
- Norris, S. (2004) *Analyzing multimodal interaction. A methodological framework*. New York: Routledge.
- Nunan, D. (1989) *Designing tasks for the communicative classroom*. Cambridge: Cambridge University Press.
- O'Halloran, K. (2004) Visual semiosis in film. In O'Halloran, K. (ed.), *Multimodal discourse analysis: Systemic functional perspective*. London: Continuum, 109–130.
- O'Halloran, K. (2011) Multimodal discourse analysis. In K. Hayland & B. Paltridge (eds.), *Companion to discourse analysis*. London & New York: Continuum, 120–137.
- Örnberg, T. (2005) Multimodality in a three-dimensional voice chat. In Allwood, J., Dorriots, B. & Nicholson, S. (eds.), *Proceedings of the 3rd Conference on Multimodal Communication, Papers in Theoretical Linguistics*. Göteborg: Göteborg University, 303–316.
- Palomeque, C. E. (2016) *Communication in a MUVE: An exploratory case study of teacher interactional devices in Second Life*. University of Barcelona, unpublished PhD. <https://www.educacion.gob.es/teseo/mostrarResult.do?ref=1246116#>
- Pellettieri, J. (2000) Negotiation in cyberspace: The role of *chatting* in the development of grammatical competence. In Warschauer, M. & Kern, R. (eds.), *Network-based language teaching: Concepts and practice*. Cambridge: Cambridge University Press, 59–86. <https://doi.org/10.1017/CBO9781139524735.006>
- Peterson, M. (2008) *An investigation of learner interaction in a MOO-based virtual environment*. University of Edinburgh, unpublished PhD. <https://core.ac.uk/download/pdf/279166.pdf>

- Pujolà, J.-T. (2002) CALLing for help: Researching language learning strategies using help facilities in a web-based multimedia program. *ReCALL*, **14**(2): 235–262. <https://doi.org/10.1017/S0958344002000423>
- Pujolà, J.-T. and Palomeque, C. (2010) Developing a multimodal transcription to account for interaction in 3D virtual worlds: The 3M method. In Bueno Alonso, J. L. (ed.), *Analizar Datos. Describir Variación*. Vigo: University of Vigo, Servizo de Publicacións, 134–145.
- Salmon, G. (2009) The future for (second) life and learning. *British Journal of Educational Technology*, **40**(3): 526–538. <https://doi.org/10.1111/j.1467-8535.2009.00967.x>
- Sweeney, P., Palomeque, C., González, D., Speck, C., Canfield, D. W., Guerrero, S. and MacKichan, P. (2011) Task design for language learning in an embodied environment. In Vincenti, G. & Braman, J. (eds.), *Teaching through multi-user virtual environments: Applying dynamic elements to the modern classroom*. Hershey, PA: Information Science Reference, 259–282.
- Tan, S., O'Halloran, K. L. and Wignell, P. (2016) Multimodal research: Addressing the complexity of multimodal environments and the challenges for CALL. *ReCALL*, **28**(3): 253–273. <https://doi.org/10.1017/S0958344016000124>
- Tudini, V. (2003) Using native speakers in chat. *Language Learning & Technology*, **7**(3): 141–159.
- Wang, A. (2015) Facilitating participation: Teacher roles in a multiuser virtual learning environment. *Language Learning & Technology*, **19**(2): 156–176.
- Warburton, S. (2009) Second Life in higher education: Assessing the potential for and the barriers to deploying virtual worlds in learning and teaching. *British Journal of Educational and Technology*, **40**(3): 414–426. <https://doi.org/10.1111/j.1467-8535.2009.00952.x>
- Warschauer, M. (ed.) (1995) *Telecollaboration in foreign language learning*. Honolulu, HI: Second Language Teaching & Curriculum Center, University of Hawaii at Manoa.
- Werry, C. C. (1996) Linguistic and interactional features of Internet relay chat. In S. C. Herring (ed.), *Computer-mediated communication: Linguistic, social and cross-cultural perspectives*. Amsterdam: John Benjamins 47–64. <https://doi.org/10.1075/pbns.39.06wer>
- Wigham, C. R. (2012) *The interplay between nonverbal and verbal interaction in synthetic worlds which supports verbal participation and production in a foreign language*. Université Blaise Pascal, unpublished PhD. <https://tel.archives-ouvertes.fr/tel-00762382>
- Williamson, R. (2007) El diseño de un corpus multimodal. *Estudios de Lingüística Aplicada*, **25**(46): 207–231.
- Yee, N., Bailenson, J. N., Urbanek, M., Chang, F. and Merget, D. (2007) The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. *CyberPsychology & Behavior*, **10**(1): 115–121. <https://doi.org/10.1089/cpb.2006.9984>

About the authors

Cristina Palomeque has a PhD in Language and Literature Education from the University of Barcelona. She has been doing CALL research for the completion of her PhD and her research interests cover foreign language learning, virtual worlds, CMC, and multimedia. She collaborates with the realTIC Research Group (<http://www.ub.edu/realtic/es/>).

Joan-Tomàs Pujolà, PhD in Applied Linguistics (University of Edinburgh), is a senior lecturer at the Department of Language Teaching in the Faculty of Education, University of Barcelona. His research focuses on CALL issues including tandem learning, material development, and new methodologies. He is the principal investigator of the realTIC Research Group (<http://www.ub.edu/realtic/es/>).