Title:
Multimodal Research: Addressing the Complexity of Multimodal Environments and the Challenges for CALL

Sabine Tan, Kay L. O'Halloran and Peter Wignell

Abstract:
Multimodality, the study of the interaction of language with other semiotic resources such as images and sound resources, has significant implications for Computer Assisted Language Learning (CALL) with regards to understanding the impact of digital environments on language teaching and learning. In this paper, we explore recent manifestations of CALL in 3-D virtual worlds, illustrated by the example of *Second Life*. The multimodal analyses of a conventional face-to-face lesson and three language learning activities in *Second Life* highlight some of the affordances and challenges presented by 3-D virtual environments. The results suggest that while multimodal resources integrate naturally to facilitate language teaching and learning in an orderly, structured and goal-orientated manner in the classroom lesson, the often uncoordinated use (or absence) of avatars' gaze, facial expression, body posture, gesture, as well as the unclear proxemics and use of space pose problems for effective communication in a 3-D virtual world. In addition, a "technology-oriented" register, alongside traditional instructional and regulative genres and registers, is introduced to help students cope with the demands of learning a language in a 3-D virtual environment. The study raises the issue of the relative effectiveness of 3-D virtual worlds for language teaching and learning. In doing so, a digital approach to multimodal research is proposed in order to address the complexity of multimodal learning environments and the various challenges for CALL.

1. Introduction

Computer Assisted Language Learning, commonly referred to as 'CALL', is "the search for and study of applications of the computer in language teaching and learning" (Levy, 1997: 1; cited in Sokolik, 2014: 19). CALL originally consisted predominately of "locally-installed software programs that offered primarily activities in which the learner consumed language, but did not produce it" (Sokolik, 2014: 19). In the 1960s and 1970s, CALL activities typically involved grammar and vocabulary drills, listening activities, and reading comprehension tasks (Levy, 2009; Yim & Warschauer, 2016). With the advancement of computer-based technologies in the 1980s and onwards, activities that required the application of multiliteracy skills and "the combination of the word and the image in the creation of multimodal texts" (Levy, 2009: 773) became popular. Today, computer mediated communication tools that require familiarity with multimodal technologies, such as audio, video and text chat, have begun to permeate a wide variety of language learning contexts. In addition, in line with the evolution of 3-D virtual environments and massively multiplayer online games (MMOGs), there has been growing interest among educators and researchers in the use and incorporation of digital games and platforms in language learning situations (Godwin-Jones, 2014).

Alongside these advances in technology, approaches to CALL have changed, which in turn have led to concordant shifts in theories and approaches to language learning and teaching in general. Following Levy (2009), Yim & Warschauer (2016: 594), conclude that the "effective integration of technology [for language learning] depends on the affordances of the particular

technology and the ways its strength and challenges can be coordinated as a pedagogical tool", Based on this perspective, this paper adopts a multimodal social semiotic approach involving the study of language and other resources (e.g. gaze, gesture, proxemics, space etc.) (e.g. Jewitt, 2014) to exploring the affordances and challenges presented by state-of-the art tools and technologies which are used for CALL, illustrated in this case by the example of 3-D virtual worlds.

In what follows, we explore multimodal approaches to CALL, with a focus on the integration of language with other semiotic resources in the communication of meaning in electronically mediated learning environments, largely from discourse analytical perspectives. We then discuss the theoretical aspects and analytical applications of a multimodal social semiotic approach for investigating the impact of digital technology on language teaching and learning, in this case for 3-D virtual worlds. The multimodal social semiotic approach is demonstrated through the analysis of a videotaped English language conversational class conducted in a conventional face-to-face setting,[1] and three short video segments that provide examples of language learning experiences in the 3-D virtual world known as *Second Life*.[2] Following this, we propose future directions for multimodal research aimed at addressing the complexity of multimodal environments and various challenges for CALL which arise in this field of study.

2. Multimodal approaches in CALL research

The significance of multimodal approaches to a variety of CALL contexts and situations has, of course, been long recognised. Unsworth (2008a), for example, discusses how the impact of information and communication technologies (ICTs) in the digital age is changing the nature of literary narratives for very young children. Unsworth (2008a) introduces multimodal frameworks to assist teachers in negotiating curricular and pedagogic approaches for developing a better understanding and appreciation of e-literacy and literacy learning using digital resources. Hampel and Hauck (2006) similarly argue that developments in computer mediated communication, which offer new ways of combining different modes and media such as text, audio and graphic, have repercussions for second language teaching and learning. They suggest that it may be useful to consider how meaning is made using the modes and media available in electronic environments, and that a multimodal approach offers a practicable framework for investigating both the limitations and the possibilities presented by new information and communication media. Like Unsworth (2008a), they suggest that a multimodal approach can contribute to an enhanced understanding of the phenomenon of new literacies, and can inform the development of enriched language learning and teaching tasks and activities in web-based environments. Hampel (2014), in turn, points out that digital tools for communication offer very different learning experiences when contrasted with traditional forms of face-to-face instruction (see also Sindoni, 2013, 2015). For online teaching to be successful, Hampel (2014) suggests an approach that is informed by a pedagogy of multiliteracies which takes in account how learners use digital media for meaning-making and for interacting and collaborating with each other, and the implications this has for language learning and teaching.

Other studies adopt a multimodal perspective for exploring how language learners utilise digital media to create multimodal texts and to interact and collaborate with each other. Several of these studies focus on the communicative behaviours of participants, and the tools

---

[1] English language conversation class: https://www.youtube.com/watch?v=AHW9AXAXgJU
[2] Second Life example English classes: https://www.youtube.com/watch?v=lB1W8oHS-W4
[2] Second Life example English classes: https://www.youtube.com/watch?v=lB1W8oHS-W4

and resources used in the context of desktop video conferencing (DVC). Berglund (2009), for example, uses multimodal interactional analysis (e.g. see Norris, 2004) to explore the influence of tool and task design on distance language learning when students are interacting with each other in a multimodal DVC environment. Satar (2013) similarly studies multimodal language learner interactions via DVC tools, with a particular focus on the use of gaze and eye-contact to facilitate the establishment of social presence. Codreanu and Combe Celik (2013), in turn, explore the effects of webcam use on the tutoring practices of experienced and trainee teachers, with a particular focus on framing (that is, the position of the subject in the frame of the screen), the degree of use of the webcam, video window utilization, the spatial context, tutors' and learners' gestures, as well as use of politeness strategies. Develotte, Guichon and Vincent (2010), likewise, focus on the utilisation of the webcam image in the context of trainee teachers learning to teach online, in combination with the use of gestures that accompany discourse and gestures used for monitoring interaction. Other multimodal studies pay close attention to the use of audio channel and text chat in video conferencing. Hampel and Stickler (2012), for example, analyse the written and spoken interactions of recorded videoconferencing sessions of foreign language learners to find out how interaction is influenced by the affordances of the environment, with particular emphasis on the discourse functions relating to social interaction, on-task negotiating meaning, off-task conversations, and technical discussions, while Stickler, Batstone, Duensing and Heins (2007) investigate the differing patterns of verbal interactions (and silences) in online and telephone tutorials of beginners' distance language courses.

Reflecting the need for new pedagogical and methodological approaches that both address and harness developments in e-learning and digital communication, other studies focus more closely on the writing process. Ciekanski and Chanier (2008), for instance, explore the ways second language learners deploy multimodal resources such as text-chat, audio, and actions performed in an integrated word-processing package, to construct a collaborative text. They argue that online writing should be perceived as both a collaborative social event as well as "a complex and procedural activity" (2008: 163), where multimodal communication is understood as being co-constructed through the deployment of semiotic resources and interactions between participants. By applying an integrative multimodal approach and methodology they aim to gain a better understanding of how participants in online environments exploit combinations of different "modes" (e.g. written, spoken language, graphic, iconic, spatial) and "modalities" (e.g. by using text-chat, conceptual map, whiteboard, word processor, audio, voting, leaving/entering a room, moving away for a moment, raising a hand, and moving between rooms and documents) to accomplish their respective learning tasks individually and collaboratively (2008: 164).

Few studies focus on the affordances of multimodal genres themselves. Ackerley and Coccetta (2007), for example, propose the use of audio and video examples of language use in context for providing non-verbal information about the environment, the participants and their moods; while Moreno Jaén and Pérez Basanta (2009) suggest that creating a database of DVD texts from various sources and contexts, could provide foreign language learners "with all semiotic modalities which may be used in oral exchanges, such as gestures and body language, facial expressions, music, etc., and make students aware of the way language interacts with other sign systems", so as to familiarize them with "the sociocultural elements of language and the subtleties of register" (Moreno Jaén & Pérez Basanta, 2009: 289).

In recent years, a vast body of research in CALL has focused on the affordances of 3-D virtual worlds and massively multiplayer online games (MOOGs) in language learning contexts.

Many of these studies (e.g. Deutschmann, Panichi & Molka-Danielsen, 2009; Neville, 2015; Peterson, 2010, 2012; Rama, Black, van Es & Warschauer, 2012) concentrate on communicative competence in social interactions, with an emphasis on the discourse features and strategies employed, such as turn-taking patterns, use of politeness involving greetings, colloquialisms, small talk, humour, leave-takings, and so forth.

Owing to the complexity of 3-D virtual worlds and MOOGs, multimodal approaches that consider the use of semiotic resources other than language are less prevalent. Zheng, Newgarden and Young's (2012) action-oriented approach, for example, demonstrates how players' meaning making and realisation of interpersonal values, such as caring for self and others, are relevant to time, location, and movements, by transcribing avatar embodied action and interactions in the virtual space of *World of Warcraft* gameplay; while Wigham and Chanier (2013) highlight the importance of nonverbal communication to overcome verbal miscommunication where direction and orientation are concerned, and how nonverbal acts (that is, proxemic and kinesic acts such as gaze, posture, gestures) were used to secure the context for deictic references to objects made in the verbal mode in the 3-D virtual world *Second Life*.

Multimodality has always been a part of language learning environments, whether deployed in 3-D virtual worlds or other ICT tools and instruments, or as illustrated, for example, in the use of gaze, gesture, posture and proxemics in face-to-face lectures and tutorials, and in the use of visual images and symbolisms in online learning materials. However, as Hampel (2014) and Sindoni (2013, 2015) point out, digital environments present very different experiences for learners than those encountered in traditional face-to-face settings. To explore some of these differences and the challenges presented by CALL tools and technologies, in this case 3-D virtual worlds, this paper adopts a multimodal approach informed by social semiotic theory, which is discussed in the following section.

3. A multimodal social semiotic approach: theory and application

The approach adopted in this paper is derived from multimodal social semiotics, based on Halliday's (1978) social semiotic theory, which perceives context and culture "as a set of semiotic systems, a set of systems of meaning, all of which interrelate" (Halliday & Hasan, 1985: 4), and which views language as but one semiotic resource among the many (e.g. image, gesture, sound, and so forth) which together constitute culture. A social semiotic approach builds upon the assumption that a text's communication function is the result of specific choices selected from a network of systems which are organised according to the functions (called "metafunctions") which the resources serve in society (e.g. Halliday & Matthiessen 2014; Martin & Rose 2007): (a) *ideational meaning,* which consists of *experiential* and *logical meaning,* to structure our experience of the world and to make logical connections in the world; (b) *interpersonal meaning* to enact social relations and to create a stance towards the world; and (c) *textual meaning* to organise experiential, logical and interpersonal meanings into coherent messages relevant to their context.

From a social semiotic viewpoint, the choices that are made in a text are not seen as the result of conscious communicative decisions "but a set of possible alternatives" (Halliday, 1994: xiv-xxvi). While the meaning potential of semiotic resources is diverse, the actual options selected in any context are conditioned by previous choices within that culture. In social semiotic theory, these configurations are conceptualised as register theory (e.g. Eggins, 2005; Martin, 1992, 2002; Martin & Rose, 2007), which is concerned with three key dimensions: (a)

field – the nature of the social activity (realised through experiential and logical choices); (b) *tenor* – the social relations which are enacted (realised through interpersonal choices); and (c) *mode* – spoken, written and visual forms of representation (realised through textual choices). The genres found in any culture are realised through the various configurations of register variables (that is, field, tenor and mode). In this case, genre is defined as "the system of staged goal-oriented social processes through which social subjects in a given culture live their lives" (Martin, 2002: 56).

Although principally informed by Halliday's social semiotic theory, multimodal approaches also draw upon other theoretical perspectives such as anthropology, sociology, critical and pragmatic discourse theory, and have been adapted and extended to the study of multimodal texts and artifacts, including visual art, paintings, architecture, sculpture, advertising, websites, toys and games (e.g. Kress & van Leeuwen, 2001, 2006; O'Toole, 2011; van Leeuwen, 1999, 2005), as well as multimodal literacy practices in educational contexts (e.g. Jewitt, 2003, 2006; Jewitt & Kress, 2003; O'Halloran, Tan & E, 2014, 2015; O'Halloran, Tan & Smith, 2016; Unsworth, 2008b). From a social semiotic viewpoint, different semiotic resources are perceived to have different affordances and constraints with regards to what can and what cannot easily be expressed or represented in a given mode (e.g. Jewitt, 2003; Kress, 2010; Kress & van Leeuwen, 2001, 2006; Machin, 2013).

In the case of traditional face-to-face learning, a range of semiotic resources such as spoken language, gaze, gesture, use of classroom space, and so forth are used (e.g. Hyland, 2009; Jewitt, 2008; Lim, O'Halloran & Podlasov, 2012). The analysis of how semiotic resources are co-deployed in the orchestration of a lesson contributes to a better understanding of the effectiveness of the various teaching strategies, tools and techniques which are used. For example, Lim, O'Halloran & Podlasov (2012) investigate how the teacher's use of classroom space impacts on the creation of a non-threatening and conducive learning environment, explaining that "different spaces in the classroom acquire specific meanings due to the typical configuration of semiotic choices in the pedagogic discourse that occurs in that space" (Lim et al., 2012: 236). Lim et al. (2012) propose four different configurations of classroom space (see CLASSROOM SPACE in Table 1a): (1) authoritative space, where the teacher's is positioned to conduct formal teaching and to provide instructions to facilitate the lesson; (2) interactional space, where the teacher moves into the students' space for personal consultation or for offering guidance and clarification on a task or activity; (3) supervisory space, where the teacher moves alongside the students' desks, or up and down the side of the classroom, primarily for the purpose of supervision during student activities; and (4) personal space, where the teacher is situated behind the teacher's desk preparing for the next stage of the lesson. The utilisation of classroom space, in combination with language and other semiotic resources, is shown to formalise the respective registers and microgenres found in classroom discourse (e.g. see Christie, 2005; Christie & Martin, 2005). For example, Christie (2005) proposes that there are two major registers in classroom discourse: the "instructional register" where the experiential meaning is dominant in terms of the lesson content, and the "regulative register" where the primary concern is interpersonal meaning and the maintenance of social relations in the classroom, as constructed through language use, in combination with other semiotic resources.

As Sindoni (2013: 2) points out, however, in digital online environments, semiotic resources are "integrated in unprecedented ways, enacting new interactional patterns and new systems of interpretation". In order to explore the contributions and interactions of linguistic and non-linguistic semiotic modes and resources (e.g. spoken language, gaze, gesture, use of space),

the utilisation of systems in a traditional face-to-face lesson (Table 1a) and those utilised differently in the communication of meaning in the context of second language learning in a 3-D virtual world (Table 1b) are investigated. These systems are not exhaustive but designed for the purpose of illustrating how the multimodal social semiotic approach can be applied to the analysis of the complex multimodal interactions in real and virtual learning environments.

Table 1a Semiotic systems used in a face-to-face conversation lesson

System/Semiotic Resource	System Choices	Description
GROUP DYNAMICS	Teacher	Refers to teacher's/instructor's speaking turns.
	Student	Refers to student's speaking turns.
GAZE	Engaged	Participant's gaze is engaged with or focused on the task/activity.
	Disengaged	Participant's gaze is disengaged.
	Indeterminable	Participant's gaze is indeterminable.
GESTURE	Interpersonal	Interpersonal gestures include regulators such as turn-taking signals (e.g. pointing/gesturing at the next speaker, or nodding to signal acceptance and understanding), and symbolic gestures or emblems (e.g. nodding to express approval).
	Experiential	Experiential gestures include illustrators that accompany a verbal message (e.g. hands forming a rectangular shape to illustrate "like a rice cake"). Experiential gestures include writing on the whiteboard.
	Textual	Textual gestures include referents (e.g. using fingers for enumerating points), and deictic gestures (e.g. pointing to give directions, or to select a specific object).
CLASSROOM SPACE	Authoritative space	Teacher is positioned in front of the teacher's desk and in the front centre of the classroom to conduct formal teaching as well as to provide instructions to facilitate the lesson.
	Interactional space	Teacher is standing alongside the students' desk or between the rows of students' desks for personal consultation where the teacher offers guidance on the task set or clarification on an earlier Instruction.
	Supervisory space	Teacher moves alongside the rows of students' desks, as well as up and down the side of the classroom, primarily for the purpose of supervision, without offering consultation to students.
	Personal space	Teacher is situated behind the teacher's desk to pack and prepare for the next stage of the lesson.

Although semiotic resources such as Group Dynamics are utilised similarly in 3-D virtual worlds, they are mediated predominately through audio channel, while resources such as Gaze and Gesture are enacted through the participants' avatars. As the analysis will show, the use of space, however, is significantly different in a 3-D virtual world. The system choices utilised in the deployment of space in the context of second language learning in the 3-D virtual world *Second Life* are displayed in Table 1b.

Table 1b Semiotic systems used in the deployment of space in a 3-D virtual world

System/Semiotic Resource	System Choices	Description
VIRTUAL SPACE	Object-oriented space	Avatars move around or stand in front of objects or interactive panels to perform a specific task.
	Disassociated space	Avatars move around in virtual space or stand motionless without performing a discernible learning task or objective.
	Detached space	Avatars' aural presence is detached from their visual presence, i.e. participants' verbal utterances can be heard over audio channel, but their avatars' facial expression is not discernible, or speaker's avatars are not present in the 3-D virtual world's visual field.

The multimodal social semiotic approach is illustrated through the analysis of a 04:47 minutes long section of a videotaped English language conversation lesson[1] in which students are acquainted with the form and use of English phrasal verbs. The analysis of the face-to-face lesson is then followed by the analysis of three short episodes from a video that offers examples of second language learning experiences in the 3-D virtual world *Second Life*[2]. The three episodes analysed are "Shopping" (05:58–07:07), an activity in which students practice manipulating their avatars while exploring virtual learning spaces in *Second Life*, "Definition Chart Race" (08:52–09:47), a learning activity which requires students to place words on a virtual definition chart to help explain the meaning of certain words, and "Affix Dominoes Game" (13:09–4:15), a game-based activity in which students practise their vocabulary skills by manipulating oversized virtual dominoes to form common words with prefixes and suffixes.

To handle the complexity of such analysis, the systems displayed in Table 1a and 1b were imported into the software tool *Multimodal Analysis Video*,[3] which is a spinoff technology developed by O'Halloran and colleagues in the Multimodal Analysis Lab in the Interactive & Digital Media Institute at the National University of Singapore. The software tool has facilities for media files to be imported and analysed using different systems, and for exporting the results to Excel for further data processing (e.g. O'Halloran, Tan & E, 2015; Tan, Smith & O'Halloran, 2015; O'Halloran, Tan & Wignell, 2016 in press).

---

[3] Multimodal Analysis Video, official website: http://multimodal-analysis.com/products/multimodal-analysis-video/

Figure 1(a) shows a snapshot of a lesson segment conducted in a face-to-face setting, as analysed in *Multimodal Analysis Video*. It illustrates the facilities for viewing the video in the player window [1] and the film strip [2], inserting time-stamped annotation nodes [3] in system strips [4], and selecting a system choice [5] from the list of available system choices [6]. All annotations are synchronised with the video player, the film strip, the sound strip [7], the time-stamped nodes in the dialog strip [8], arranged in separate tabs for each speaker [9], and the corresponding verbal transcriptions in the transcription window [10].
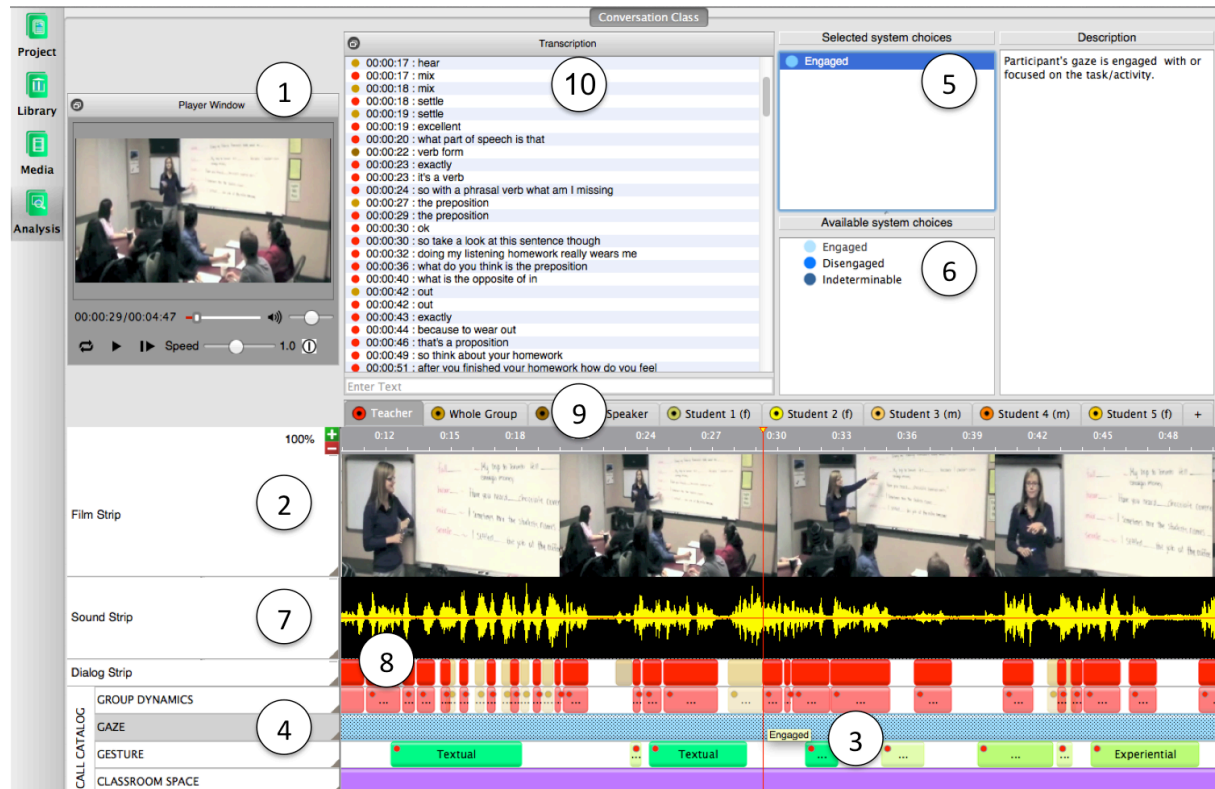
Figure 1(a). Analysis of face-to-face lesson performed in Multimodal Analysis Video. Notation: Player window [1], film strip [2], time-stamped annotations [3], system strips [4], selected system choice [5], list of available system choices [6], sound strip [7], dialog strip [8], speaker tabs [9]; transcription of verbal utterances [10]

The software also provides the option to visualise the interaction of different semiotic systems in the form of state transition diagrams (see Figure 1(b)), whereby a 'state' (denoted by circles) [1] represents the system choices [2] that have been utilised for a particular system or a combination of semiotic systems [3], displayed in terms of total video time. The lines between the states [4] represent the movement (i.e. 'transitions') between individual states as the lesson unfolds. In essence, the state transition diagram provides visual access to how combinations of semiotic choices are deployed in the lesson in terms of relative time, and the nature of the transitions which occur between those combinations of choices during the lesson.
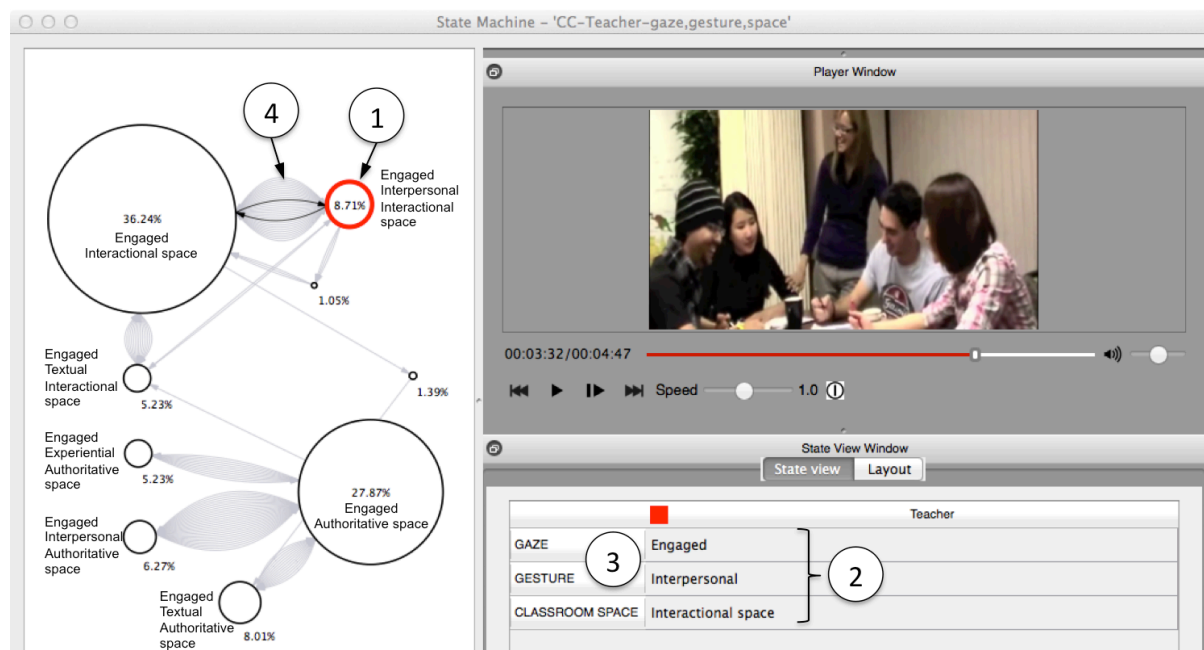
Figure 1(b). Screenshot of state transition diagram for the systems of Gaze, Gesture and Classroom Space as deployed by the teacher in the videotaped face-to-face lesson. Notation: State [1], combination of system choices in a state [2], selected systems in the analysis [3], transitions between states [4]

4. Analysis and Discussion

In what follows, the use of language and other semiotic systems (as displayed in Table 1a and 1b) are analysed in the face-to-face lesson and the excerpts of learning activities in *Second Life*, in order to investigate how the utilisation of semiotic systems in a traditional face-to-face lesson differs from those utilised in a 3-D virtual world, and how the two different learning environments impact on second language teaching and learning.
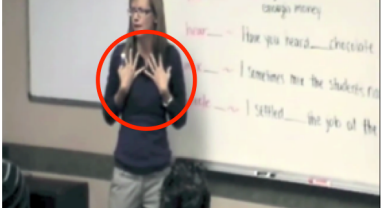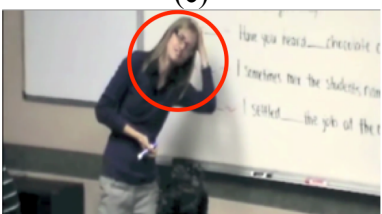
4.1 Face-to-face lesson
The analysis of the videotaped face-to-face lesson shows that the teacher utilises only two of the pedagogical spaces proposed by Lim et al. (2012), namely the authoritative and the interactional space. Indeed, the lesson is structurally divided almost equally into two distinct learning spaces. For the first half of the videotaped lesson, the teacher is positioned in authoritative space (02:15 minutes or 47% in terms of total video time), where the task is to acquaint students with the form and meaning of common English phrasal verbs, whilst in the second half of the video she utilises interactional space (02:27 minutes or 52% in terms of total video time), where students are tasked with the practice of applying some of the phrasal words learnt in the first half of the lesson in group work. Through the entire duration of the videotaped lesson, the teacher's (as well as the students') gaze is engaged with the learning task. In authoritative space, her gaze is either directed at the whiteboard while she interacts with it to explicate and scaffold the learning task, or directed at the students, who sit at their desks facing her and the whiteboard. In interactional space, both the teacher's and the students' gaze is engaged with the activity and each other.

Regardless of whether she is conducting formal teaching or offering guidance and clarification on the task at hand, the teacher makes extensive use of gesture to create (a) textual meanings (13.24% in terms of total video time) to orchestrate other semiotic choices (e.g. "so take a look at this sentence"); (b) experiential meanings (6.27% in terms of total

video time) to construct experience (e.g. "how does your body feel"); and (c) interpersonal meanings (14.98% in terms of total video time) to enact social relations and create an stance towards happenings in the world (e.g. "after you finish your homework, how do you feel?"). Table 2 provides examples of the types of gestures deployed in combination with verbal utterances in the face-to-face lesson. The use of gesture in combination with language and space is further explored in detail below.

Table 2 Examples of gestures used in face-to-face lesson

Type of gesture	Example	Verbal utterance
Textual	(a)	Teacher: 00:30: so take a look at this sentence though
	(b)	Student 4 (male): 03:12: oh it's just three 03:13: oh between three
Experiential	(c)	Teacher: 00:56: how does your body feel
	(d)	Teacher: 00:58: how does your mind feel
	(e)	Student 5 (female): 03:47: it's just like a rice cake
Interpersonal	(f)	Teacher: 00:51: after you finished your homework how do you feel

		Teacher: 03:32: ok Emily your turn which food
	(g)	
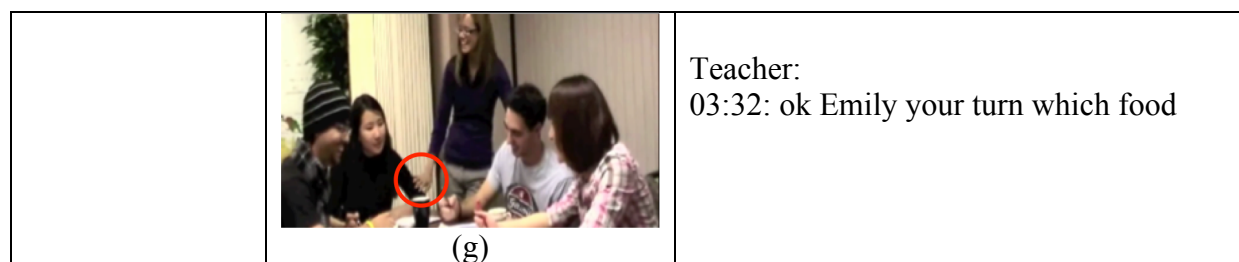
As displayed in the transitional diagram in Figure 1(b), the use and meaning potential of gestures differs according to the learning task and use of pedagogic space. That is, owing to the instructional nature of the register used for teaching content (e.g. Christie 2005; Christie & Martin, 2005), the majority of the teacher's gestures in authoritative space (8.01% in terms of total video time) are deployed for realising textual meaning. Consisting of referential and deictic gestures, these are used by the teacher to accentuate specific nuances of the learning task, for example, by sequentially pointing to the phrasal words written on the whiteboard while articulating them verbally (e.g. "wear, fall, hear, mix, settle"), or by pointing to specific aspects of the task at hand (e.g. "so take a look at this sentence though") (see Table 2, Example (a)). To explicate the meaning of certain words and concepts, the teacher also deploys gestures that realise experiential meaning (5.23% in terms of total video time). For example, to elucidate the concept that "to wear out means to feel tired", she touches both her hands to her chest when asking the question "how does your body feel" (Table 2, Example (c)), and touches her hand to her inclined head while asking the question "how does your mind feel" (Table 2, Example (d)). In addition, she deploys a range of gestures to realise interpersonal meaning (6.27% in terms of total video time), such as spreading her arms with palms facing upwards to elicit a choral response from the students (Table 2, Example (f)). Not surprisingly perhaps, interpersonal gestures predominate in interactional space (8.71% in terms of total video time), consisting mostly of regulators for selecting speakers either by pointing, gesturing or intimately touching the next speaker on the arm (Table 2, Example (g)), as well as symbolic gestures or emblems, such as nodding for signalling acceptance and understanding or to express approval. While students make use of gesture less frequently than the teacher in the videotaped lesson, they do deploy experiential gestures, for example, by forming a rectangular shape with both hands to illustrate that the type of food selected by the student as an example of a food group is "just like a rice cake" (Example (e)), as well as textual gestures, e.g. for clarifying and selecting one example from a range of food groups by pointing to the written instructions lying in front them) (Table 2, Example (g)).

The analysis shows that the group dynamics of the face-to-face lesson are very much teacher-led and directed (e.g. the teacher speaks for 51.22% in terms of total video time and the students speak collectively for 32.76% of the time). Any misunderstandings about the learning task are quickly resolved through the non-ambiguous interaction of speech and non-verbal cues. For instance, the teacher's prompt "after you finished your homework how do you feel" initially elicits the (contextually incorrect) choral response "happy" from the students. By adapting the prompt to "how does your body feel" and "how does your mind feel", performed in combination with gesture (see Table 2, Examples (c) and (d)), the teacher effortlessly elicits the contextually appropriate response of "tired". Lim et al. (2012: 247) introduce the concept of "structured informality" to describe to the ways in which non-threatening and conducive learning environments are created in the classroom, where students feel comfortable and are encouraged to respond and speak up, while the lesson unfolds in an orderly, structured and goal-oriented manner, reinforced by the use of well-coordinated semiotic resources.

4.2 Second Life
The analysis of the excerpts of different language learning contexts as experienced in the 3-D virtual world *Second Life*, performed with *Multimodal Analysis Video* software, shows that the multimodal approaches developed for investigating conventional pedagogical genres and face-to-face interactions need to be sufficiently adapted to facilitate an examination of the affordances and challenges presented by 3-D virtual worlds in the context of language learning. One of the affordances of 3-D virtual worlds as CALL tools is that students are no longer bound by the confines of a material world, but are free to explore and traverse these virtual spaces in ways that are quite different from those experienced in a traditional classroom setting. While offering new, immersive game-based experiences to language learners, the following analysis will show that navigating a 3-D virtual world in the pursuit of pedagogical tasks and activities also presents new challenges.

The utilisation of space in *Second Life*, in particular, is unlike any experience that students would encounter in a traditional learning environment. Instead of the pedagogical spaces described by Lim et al. (2012), the analysis shows that in the virtual world avatars are either positioned in front of objects or interactive panels to perform a specific task (which we term 'object-oriented' space), or they move or stand around in virtual space without performing a discernible learning task or objective (termed 'disassociated space'), or the avatars' aural presence is detached from their visual presence (termed 'detached space'), that is, participants' voices can be heard over the audio channel, but their avatars are either not present in the visual field, or their avatars' facial expressions are not discernible (e.g. see Table 1b). Moreover, in contrast to real-life settings where teachers and students are likely to occupy the same learning spaces, the use of space in a virtual learning environment (at least by novice users) can appear arbitrary, unnatural, and uncoordinated, as evidenced by the analysis of the view of the 3-D world presented in the video excerpts analysed below.

For example, the analysis of the featured activity "Shopping" (Figure 2), which is designed to ease students into manipulating (and learning through) their avatars in *Second Life*, shows a group of students starting their learning activity by positioning their avatars in object-oriented space in front of a display panel from which they select items of clothing, most likely in response to verbal cues provided by their teacher. Meanwhile, other avatars stand motionless or are shown milling around in the background. Only a few avatars perform gestures that are coordinated with their interactions. The avatar belonging to Student 1, for example, who is pictured facing the display panel, can be observed performing interpersonal and textual gestures (e.g. nodding and pointing) in response to the instructor's question "Sherry have you found something?" (see Figure 2). The scene just described is displayed in the Player Window [1] and first visual frame of the Film Strip, marked [2]. The range of semiotic systems utilised is denoted by the annotation nodes in the strips labelled Gaze, Gesture, Virtual Space, marked [3]. Most avatars in this exercise, however, move around unnaturally, with their hands stiffly at their sides. Moreover, the avatars' disengaged or indeterminable gaze and ambiguous proxemics can make it difficult to follow who is being addressed or who is talking when not nominated explicitly by the teacher. For example, in response to the teacher's command to a student to "choose (someone) who is in front of you and tell me what she is wearing", a student responds by describing an avatar who "wears a long blue dress that is very gorgerous [sic] and very beautiful, and her back is naked". While the student is still speaking, the avatar wearing the blue dress wanders off into the far distance. The avatars' uncoordinated movement and positioning, coupled with their disengaged or indeterminable gaze, has the effect that the activity takes place in disassociated, or even detached space for some students

(see overlapping nodes in the strips labelled 'Gaze' and 'Virtual Space' in Figure 2). This effect could be exacerbated by the individual camera point of view which can be adopted by the different participants in *Second Life*. In this exercise, avatars spend only 9.59% of the total duration of this video segment in object-oriented space.
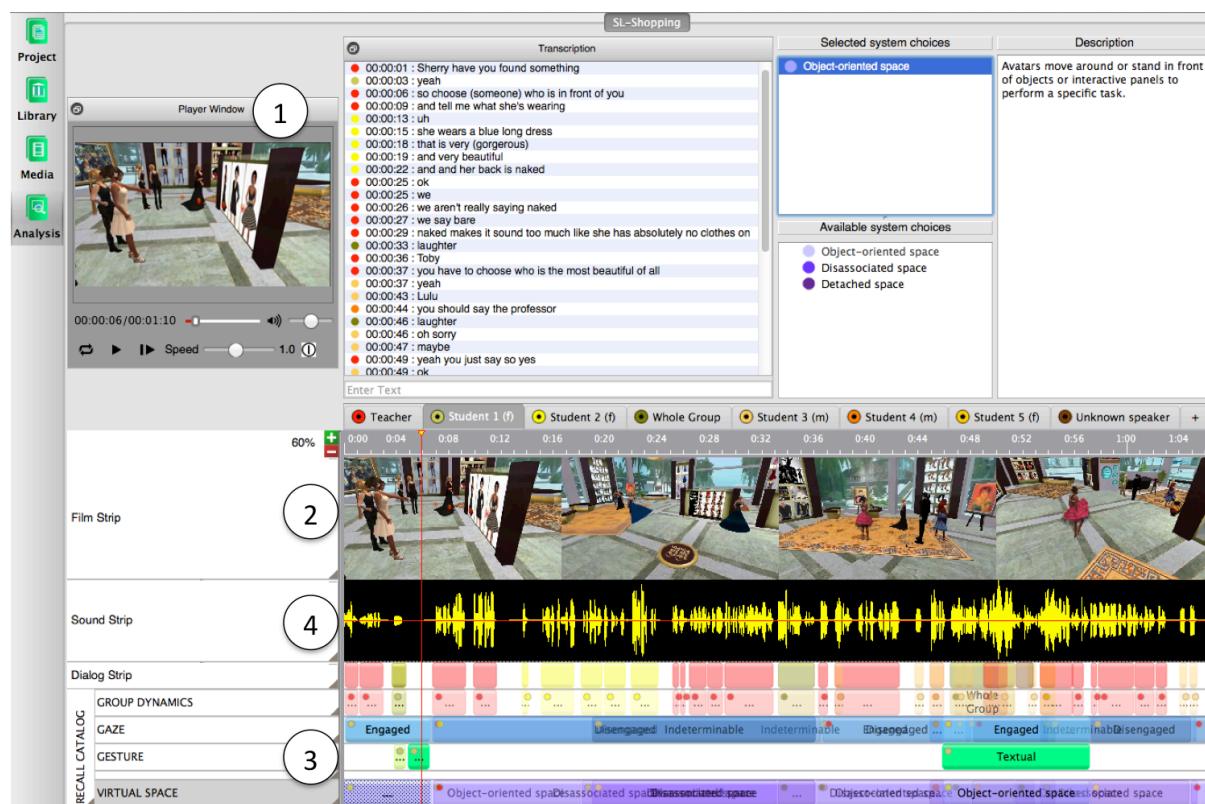

Figure 2. "Shopping". Notation: Player window [1], film strip [2], system strips for the systems of Gaze, Gesture, and Virtual Space [3], sound strip [4]

In terms of group dynamics, the verbal interactions are mostly teacher-led and directed (e.g. the teacher speaks for 42.25% of the time in terms of the total duration of the video sequence, while the students collectively speak for 26.77% of the time). The interaction is orchestrated predominately through aural mode by means of directives addressed to particular students (e.g. Teacher: "Toby, you have to choose who is the most beautiful of all"). There is some (limited) spontaneous group interaction arising from humour in reaction to a student's suggestion "you should say the professor" in response to the teacher's command to Toby, characterised by laughter and overlapping speech. On the whole, however, the dialogic interactions in this activity take place sequentially between the teacher and a student, one at a time, often with a time lag between each interaction. A lack of natural rhythm, caused no doubt by technological problems, such as unclear, disrupted or lagging audio, lends the interaction an unnatural and spasmodic feel (as reflected by the short, sharp bursts of jagged sound waves in the Sound Strip [4] in Figure 2).

The group dynamics in the second excerpt analysed, the "Definition Chart Race" (Figure 3), a game-based learning activity in which students are tasked with placing words on a virtual definition chart to help explain the meaning of a certain word or concept, are similar. The activity is again teacher directed, and orchestrated through the teacher's verbal directives and explanations using mainstream instructional register for teaching content, for example, by asking focused questions such as "what word will sour help explain" to elicit the (correct) answer "a taste" from the student. The activity, however, then shifts into technology-oriented

register with the teacher directing the student to "move" her word into the "lemon column in the taste row", to which the student responds by virtually manipulating the label with word "sour" up on the definition chart without physically interacting with it, just as one would move a word by manipulating a mouse on a computer screen. Unfortunately, the word ends up getting stuck in the wrong place on the definition chart (see Player Window [1] and third visual frame in the Film Strip [2] in Figure 3). The student's indirectly phrased requests for help in solving a problem posed by technology (Student 1: "do you know where is the refresh, I just not sure", "I don't know to refresh is what") fail to elicit the desired response. One student interjects to offer assistance, but confuses the request for help with the technology-oriented task of refreshing the screen with the instructional register and proceeds to painstakingly explain the meaning of the word "refresh" (Student 2: "to refresh means …"). The misunderstanding takes several speech turns to resolve, with no input from the teacher or the other students (e.g. see Transcription [3] in Figure 3). While they are virtually present through their avatars, they do not participate in the unfolding verbal exchange between the two students. Most avatars are positioned facing the definition chart, with some performing textual movements (e.g. pointing, gesturing at the chart), whereas others exhibit kinetic movements that do not seem to directly contribute (e.g. flying around). While their body posture and orientation suggests that they are at least engaged with the activity, other avatars stand completely motionless.
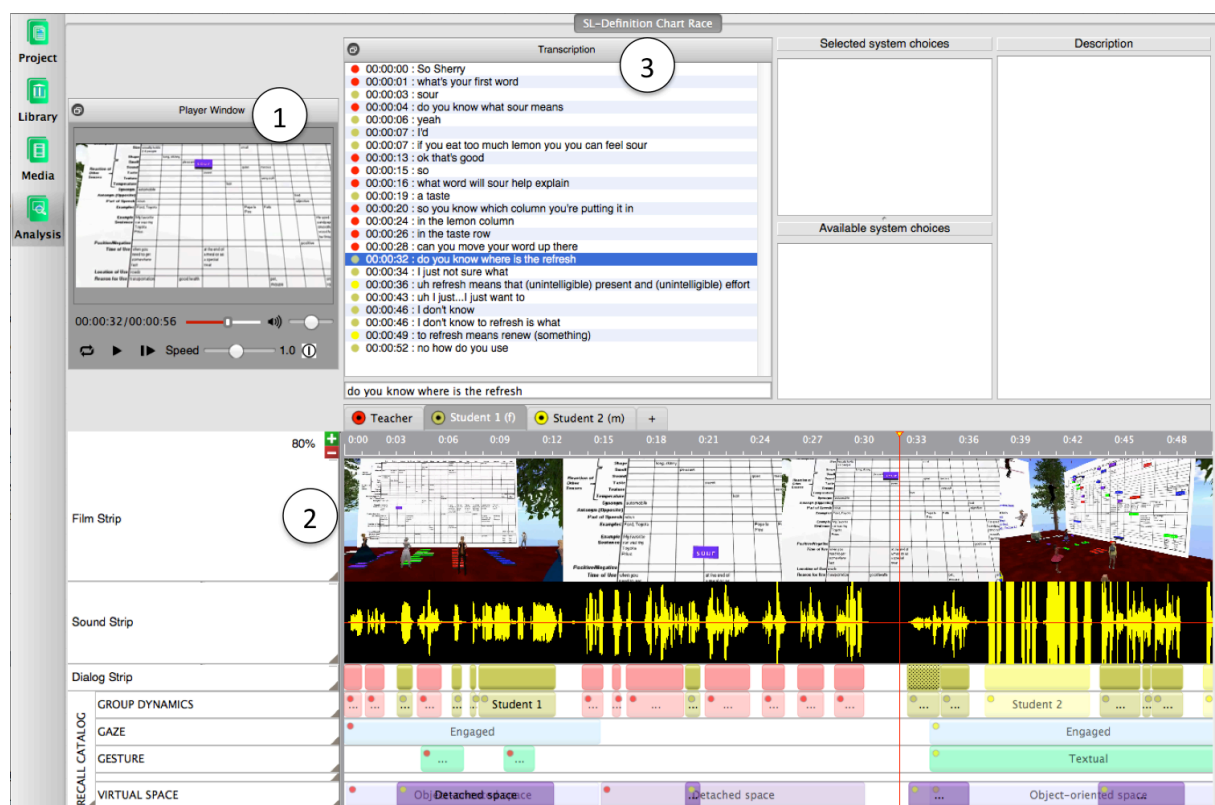

Figure 3. "Definition Chart Race". Notation: Player window [1], film strip [2], transcription [3]

The third activity analysed is the "Affix Dominoes Game" (Figure 4), where students compete in pairs to practise their vocabulary skills by playing a floor-game with oversized dominoes whose surfaces have been replaced with common word groups, prefixes and suffixes. The group dynamics in this activity are slightly different from the previous two activities in that the verbal exchanges in this exercise are initiated predominately by the students in the form of statements, questions and commands addressed to the teacher and other students. Overall,

students' dialogic contributions increased significantly in this exercise and account for 52.93% of the total duration of this video segment. The teacher speaks considerably less in this segment (17.65% in terms of the video segment's duration), but still maintains control over the exercise by correcting students' errors and providing feedback (instructional register) and by monitoring and maintaining classroom interactions (regulative register). The utilisation of virtual space also appears to be more coordinated in this exercise in that students' avatars spend more time in object-oriented space (17.65% in terms of video segment duration), with their gaze generally directed at the dominoes (14.71% in terms of video segment duration). Wayward avatars spending time in disassociated space are quickly re-orientated towards object-oriented space by the teachers' mood metaphor commands (e.g. Teacher: "ok guys your turn"), to which two students react by complying verbally ("ok, coming") while their avatars immediately follow the teacher's avatar to the designated space to interact with the dominoes game (see scene displayed in the Player Window [1] and the first visual frame in the Film Strip [2], and the system strip labelled Virtual Space [3] in Figure 4).
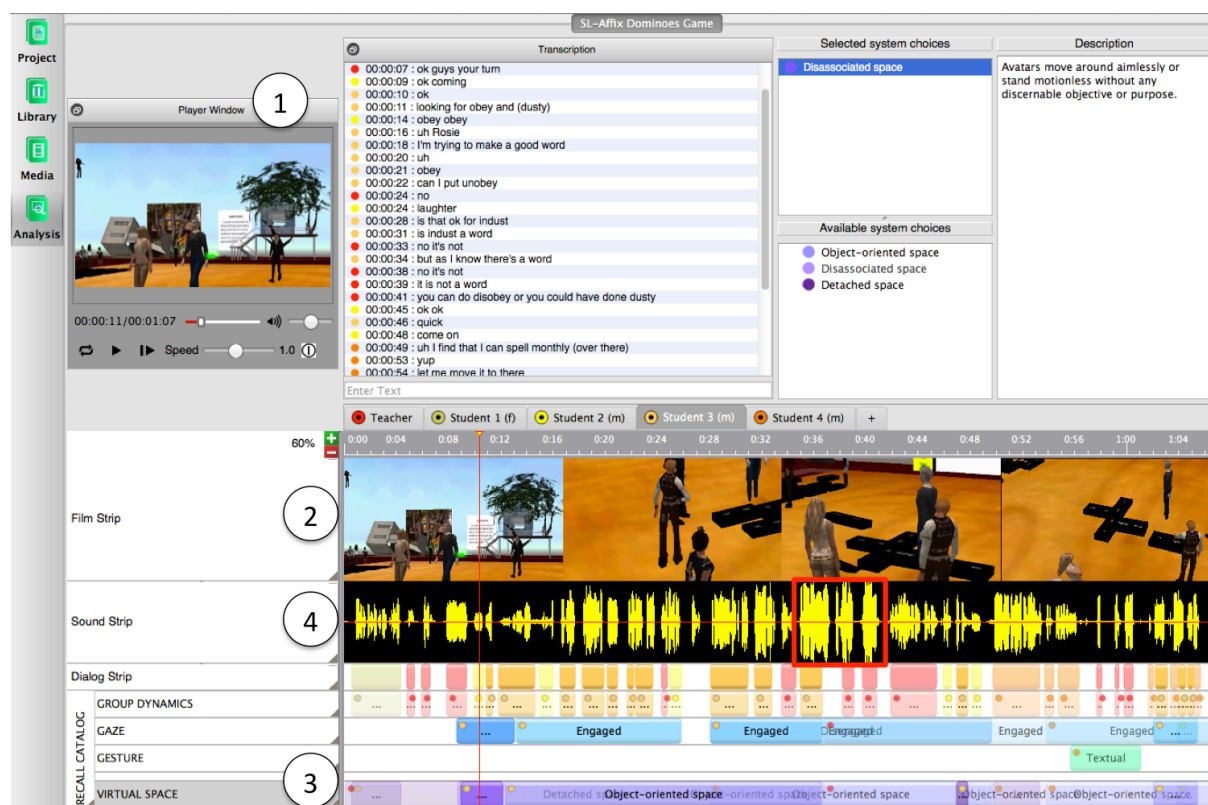

Figure 4. "Affix Dominoes Game". Notation: Player window [1], film strip [2], system strip for Virtual Space [3], sound strip [4]

Although the competitive nature of game-based activities can place challenges on the traditional hierarchical nature of teacher-student interactions, the relative anonymity afforded by a 3-D virtual world can provide opportunities for students to take control over their own learning by exploring verbal strategies for arguing their point in a non-threatening environment. In this video sequence, for instance, we find one student verbally challenging the teacher about the combinational possibilities of affixes with common word stems (see excerpt of transcription (a) in Table 3). The verbal exchange is marked clearly by a change in pitch and tone (indicated by the spikes in the sound waves, outlined in red in the Sound Strip [4], underneath the third visual frame in the film strip in Figure 4), but is not manifest in the avatars' movement, gestures, facial expression, or body posture, as they stand motionless with

their hands at their sides during this scene. Avatars' actions and gestures are generally of a textual nature, and coordinated (to some extent) with other semiotic resources, such as language, where the movement of objects (i.e. dominoes) is concerned. One student, for example, can be seen performing textual gestures (e.g. pointing) when selecting and manipulating a domino piece (Student 4: "let me move it to there" (third line of transcription (b) in Table 3; see also last visual frame in the film strip [2] in Figure 4). However, while the vocabulary puzzle was solved in a single utterance using verbal mode (first line of transcription (b) in Table 3), problems posed by technology are again shown to interfere with the completion of the language learning task, as the student's avatar appears to have difficulties in joining the domino pieces together, requiring the intervention of another student (see excerpt of transcription (b) in Table 3).

Table 3 Excerpts of verbal exchanges in "Affix Dominoes Game"

Scene	Excerpt of Transcription
(a) Challenging the teacher	Student 3 00:00:18 : I'm trying to make a good word Student 3 00:00:20 : uh Student 3 00:00:21 : obey Student 3 00:00:22 : can I put unobey Teacher 00:00:24 : no 00:00:24 : laughter Student 3 00:00:28 : is that ok for indust Student 3 00:00:31 : is indust a word Teacher 00:00:33 : no it's not Student 3 00:00:34 : but as I know there's a word Teacher 00:00:38 : no it's not Teacher 00:00:39 : it is not a word Teacher 00:00:41 : you can do disobey or you could have done dusty
(b) Interference of technology	Student 4 00:00:49 : uh I find that I can spell monthly (over there) Student 4 00:00:53 : yup Student 4 00:00:54 : let me move it to there Teacher 00:00:57 : yeah Teacher 00:00:59 : good Teacher 00:00:59 : a little more Student 4 00:01:01 : yeah Student 3 00:01:02 : let me move it Student 4 00:01:03 : alright Student 3 00:01:04 : put it down Student 3 00:01:04 : put it down Student 4 00:01:05 : alright

Applying a multimodal social semiotic approach, which views language and other semiotic resources as serving particular communicative functions in society (i.e. *experiential* meaning for structuring our experience of the world; *interpersonal* meaning for enacting social relations and for creating a stance towards the happenings in this world; and *textual* meaning for organising experiential and interpersonal meanings into coherent messages relevant to their context), the above analysis has shown that the orchestration of meaning in a face-to-face lesson is very different from the language learning experiences afforded by a 3-D virtual world such as *Second Life*, in particular with regard to the coordinated use of language, gaze, gesture, and space.

On the one hand, the analysis has shown that the coordinated interplay of instructional registers (for communicating experiential meaning in terms of lesson content) and regulative registers (for communicating interpersonal meaning), integrated with the use of gaze and gesture, and orchestrated in the utilisation of classroom space, enabled the face-to-face lesson

to unfold in an orderly, structured and goal-oriented manner while providing, at the same time, a conducive and participative learning environment to students.

On the other hand, the analysis indicates that while 3-D virtual worlds can provide a safe and conducive environment for students, by encouraging them to speak more (and perhaps more uninhibited) than they would in a classroom based environment, learning a language in a virtual environment can present additional challenges. For instance, the often uncoordinated use (or absence) of avatars' gaze, facial expression, body posture, gesture (in particular the lack of interpersonal gestures to enact social relations and experiential gestures to help structure the content of the lesson), coupled with the unclear proxemics and use space can pose problems for effective communication in a 3-D virtual world. Moreover, a "technology-oriented" register (for dealing with problems posed by technology) is introduced alongside traditional instructional and regulative genres and registers in order to help students cope with the demands of learning a language in a 3-D virtual world. As the analysis of the three learning activities in *Second Life* suggests, the successful incorporation of the technology-oriented register into traditional teaching and learning strategies can be problematic, particularly for novice learners. Interference of technology (e.g. frozen screens, unclear or lagging audio), coupled with learners' inexperience and unclear task design, can lead to communication breakdown, which has implications for effective language learning.

5. Conclusion and future directions for multimodal research in CALL

In the above sections we have explored multimodal approaches to research in CALL, with a focus on the integration of language with other semiotic resources in the communication of meaning in electronically mediated learning environments. We then introduced the theoretical aspects and analytical applications of a multimodal social semiotic approach for investigating the differing contributions and interactions of multimodal resources in the context of learning a second language in a traditional classroom-based environment and those encountered in a 3-D virtual world. The approach was demonstrated through the multimodal analysis of an English language lesson conducted in a conventional face-to-face setting, followed by the analysis of three examples of language learning experiences in *Second Life*.

3-D virtual worlds have generated a lot of interest with regard to their use as a CALL tool (e.g. Canto, Jauregi & van den Bergh, 2013; Deutschmann, Panichi & Molka-Danielsen, 2009; Jauregi, Canto, de Graaff, Koenraad & Moonen, 2011; Neville, 2015; Peterson, 2010, 2012; Rama, Black, van Es & Warschauer, 2012; Wigham & Chanier, 2013; Zheng, Newgarden & Young, 2012), and are much valued for the (almost infinite) variety of different contexts that can be created and explored, and the opportunities they offer for learning a language in a non-threatening environment, especially for geographically dispersed students. However, in order to provide students with affordances that approximate those found in conventional face-to-face settings (as far as the complex orchestration of semiotic modes and resources is concerned), 3-D virtual worlds and massively multiplayer online games (MOOGs) would need to evolve technically much further.

The affordances, challenges and complexities of multimodal discourse genres, as outlined above, have implications not only for language learning and teaching, but also for future research directions in CALL. In order to appreciate the demands of continuously evolving multimodal tools and techniques, which continue to transform and impact traditional teaching and learning practices, greater awareness needs to be generated about the communicative functions of individual semiotic resources (e.g. spoken and written language, gaze, gesture,

and so forth) and the multiple ways they combine and interact in digitally mediated environments. The multimodal social semiotic approach as a theoretical tool is of evident interest to researchers, educators, and even developers, for this purpose, resulting in a better understanding of the effectiveness of the CALL tools and techniques, as evidenced by earlier work and this study.

While the multimodal social approach provides a comprehensive theoretical platform for approaching the use of digital platforms for language learning, a major challenge is actually using that framework to analyse language learning in multimodal environments, given the evident complexity of the interactions which are involved. This problem is overcome in this study by using purpose-built multimodal analysis software, but even then, the analysis is limited in terms of scope and focused on the orchestration of only a few select semiotic resources that may be deployed in the context of language learning in a 3-D environment. Moreover, the data set in the current study is limited to extracts from one classroom lesson and three learning activities in a 3-D virtual world. In order to develop an evidence-based approach to CALL research to support research findings, further digital tools and techniques are required in order to analyse multimodal interactions in larger datasets. That is, through the development of new digital tools and techniques for multimodal analysis which are underpinned by multimodal social semiotic theory, the impact of digital technologies on teaching and learning language can be assessed, with a view to developing new innovative platforms to ensure effective practices for CALL.

Websites
1   https://www.youtube.com/watch?v=AHW9AXAXgJU
2   https://www.youtube.com/watch?v=lB1W8oHS-W4
3   http://multimodal-analysis.com/products/multimodal-analysis-video/

References

Ackerley, K. & Coccetta, F. (2007) Enriching language learning through a multimedia corpus. *ReCALL*, **19**(3): 351–370.
Berglund, T. Ö. (2009) Multimodal student interaction online: an ecological perspective. *ReCALL*, **21**(2): 186–205.
Canto, S., Jauregi, K., & van den Bergh, H. (2013) Integrating cross-cultural interaction through video-communication and virtual worlds in foreign language teaching programs: is there an added value? *ReCALL*, **25**(1): 105–121.
Christie, F. (2005) *Classroom Discourse Analysis: A Functional Perspective*. London & New York: Continnum.
Christie, F. & Martin, J. R. (2005) *Genre and Institutions: Social Processes in the Workplace and School*. London & New York: Continuum.
Ciekanski, M. & Chanier, T. (2008) Developing online multimodal verbal communication to enhance the writing process in an audio-graphic conferencing environment. *ReCALL*, **20**(2): 162–182.
Codreanu, T. & Combe Celik, C. (2013) Effects of webcams on multimodal interactive learning. *ReCALL*, **25**(1): 30–47.
Deutschmann, M., Panichi, L., & Molka-Danielsen, J. (2009) Designing oral participation in Second Life – a comparative study of two language proficiency courses. *ReCALL*, **21**(2): 206–226.
Develotte, C., Guichon, N., & Vincent, C. (2010) The use of the webcam for teaching a foreign language in a desktop videoconferencing environment. *ReCALL*, **22**(3): 293–312.

Eggins, S. (2005) *An Introduction to Systemic Functional Linguistics* (2nd edition). London: Pinter Publishers.

Godwin-Jones, R. (2014) Games in language learning: opportunities and challenges. *Language Learning & Technology* **18**(2): 9–19.

Halliday, M. A. K. (1978) *Language as Social Semiotic*. London: Edward Arnold.

Halliday, M. A. K. (1994) *An Introduction to Functional Grammar* (2nd edition). London: Edward Arnold.

Halliday, M. A. K. & Hasan, R. (1985) *Language, Context and Text: Aspects of Language in a Social-Semiotic Perspective*. Waurn Ponds, Victoria: Deakin University.

Halliday, M.A. K. & Matthiessen, C. M. I. M. (2014). *Halliday's Introduction to Functional Grammar*. London & New York: Routledge.

Hampel, R. (2014) Making meaning online: computer-mediated communication for language learning. In Peti Stantić, A. & Stanojević, M. M. (eds.), *Language as Information. Proceedings from the CALS Conference 2012*. Frankfurt am Main: Peter Lang, 89–106.

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

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

Hyland, K. (2009) *Academic Discourse: English In A Global Context*. London: Continuum International Publishing.

Jauregi, K., Canto, S., de Graaff, R., Koenraad, T., & Moonen, M. (2011) Verbal interaction in *Second Life*: towards a pedagogic framework for task design. *Computer Assisted Language Learning*, **24**(1): 77–101.

Jewitt, C. (ed.) (2014) *The Routledge Handbook of Multimodal Analysis*. London: Routledge.

Jewitt, C. (2003) Re-thinking assessment: multimodality, literacy and computer-mediated learning. *Assessment in Education: Principles, Policy & Practice*, **10**(1): 83–102.

Jewitt, C. (2006) *Technology, Literacy and Learning: A Multimodal Approach.* London: Routledge.

Jewitt, C. (2008) Multimodality and literacy in school classrooms. *Review of Research in Education*, **32**(1): 241–267.

Jewitt, C. & Kress, G. (eds.) (2003) *Multimodal Literacy*. New York: Peter Lang.

Kress, G. (2010) *Multimodality: A Social Semiotic Approach to Contemporary Communication*. New York: Routledge.

Kress, G. & van Leeuwen, T. (2001) *Multimodal Discourse: The Modes and Media of Contemporary Communication Discourse*. London: Arnold.

Kress, G. & van Leeuwen, T. (2006) *Reading Images: The Grammar of Visual Design* (2nd edition). London: Routledge.

Levy, M. (1997) *CALL: Context and Conceptualisation*. Oxford: Oxford University Press.

Levy, M. (2009) Technologies in use for second language learning. *The Modern Language Journal*, 93, Focus Issue: 769–782.

Lim, F. V., O'Halloran, K. L., & Podlasov, A. (2012) Spatial pedagogy: mapping meanings in the use of classroom space. *Cambridge Journal of Education,* **42**(2): 235–251.

Machin, D. (2013) What is multimodal critical discourse studies? *Critical Discourse Studies* (10)4: 347–355.

Martin, J. R. (1992) *English Text: System and Structure*. Amsterdam: Benjamins.

Martin, J. R. (2002) Meaning beyond the clause: SFL perspectives. *Annual Review of Applied Linguistics* 22: 52–74.

Martin, J. R. & Rose, D. (2007) *Working with Discourse: Meaning beyond the Clause* (2nd edition). London: Continuum.

Moreno Jaén, M. & Pérez Basanta, C. (2009) Developing conversational competence through language awareness and multimodality: the use of DVDs. *ReCALL*, **21**(3): 283–301.

Neville, D. O. (2015) The story in the mind: the effect of 3D gameplay on the structuring of written L2 narratives. *ReCALL*, **27**(1): 21–37.

Norris, S. (2004) *Analyzing Multimodal Interaction: A Methodological Framework*. London & New York: Routledge.

O'Halloran, K. L., Tan, S., & E, M. K. L (2014) Multimodal pragmatics. In Schneider, K. & Barron, A. (eds.), *Pragmatics of Discourse*. Berlin: De Gruyter Mouton, 239–268.

O'Halloran, K. L., Tan, S.**, &** E, M. K. L. (2015) Multimodal analysis for critical thinking. *Learning, Media and Technology*. DOI: 10.1080/17439884.2016.1101003.

O'Halloran, K. L., Tan, S., & Smith, B. A. (2016) Multimodal approaches to English for academic purposes. In Hyland, K. & Shaw, P. (eds.), *The Routledge Handbook of English for Academic Purposes*. London & New York: Routledge, 257–270.

O'Halloran, K. L., Tan, S., & Wignell, P. (2016 in press) Inter-semiotic translation as resemiotization: a multimodal perspective. In Fontanille, J., Sonzogni, M., & Troqe, R. (eds.), *Signata. Special Issue on Translating: Signs, Texts, Practices*.

O'Toole, M. (2011) *The Language of Displayed Art* (2nd edition). London/New York: Routledge.

Peterson, M. (2010) Learner participation patterns and strategy use in Second Life: an exploratory case study. *ReCALL*, **22**(3): 273–292.

Peterson, M. (2012) Learner interaction in a massively multiplayer online role playing game (MMORPG): a sociocultural discourse analysis. *ReCALL*, **24**(3): 361–380.

Rama, P. S., Black, R. W., van Es, E., & Warschauer, M. (2012) Affordances for second language learning in World of Warcraft. *ReCALL*, **24**(3): 322–338.

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

Sindoni, M. G. (2013) *Spoken and Written Discourse in Online Interactions. A Multimodal Approach*. London & New York: Routledge.

Sindoni, M. G. (2015) Teaching English multimodally. The use of new travel websites in EFL environments. *Teaching English with Technology*, **15**(2): 19–38.

Sokolik, M. (2014) What constitutes an effective Language MOOC? In Martín-Monje E. & Bárcena Madera, E. (eds.), *Language MOOCs: Providing Learning, Transcending Boundaries*. Berlin: De Gruyter Open, 16–32.

Stickler, U., Batstone, C., Duensing, A., & Heins, B (2007) Distant classmates: speech and silence in online and telephone language tutorials. *European Journal of Open, Distance and e-Learning*, **2007**(2). http://oro.open.ac.uk/32919/1/Stickler_Batstone_Duensing_Heins.html

Tan, S., Smith, B. A., & O'Halloran, K. L. (2015) Online leadership discourse in higher education: A digital multimodal discourse perspective. *Discourse & Communication*, **9**(5): 559–584.

Unsworth, L. (2008a) Multiliteracies, E-literature and English teaching. *Language and Education*, **22**(1): 62–75.

Unsworth, L. (ed.) (2008b) *Multimodal Semiotics: Functional Analysis in Contexts of Education*. New York: Continuum.

van Leeuwen, T. (1999) *Speech, Music, Sound*. Houndsmills: Macmillan Press.

van Leeuwen, T. (2005) *Introducing Social Semiotics*. London: Routledge.

Wigham, C. R. & Chanier, T. (2013) A study of verbal and nonverbal communication in Second Life – the ARCHI21 experience. *ReCALL*, **25**(1): 63–84.

Yim, S. & Warschauer, M. (2016) CALL and electronic media. In Hyland, K. & Shaw, P. (eds.), *The Routledge Handbook of English for Academic Purposes*. London & New York: Routledge, 592–607.

Zheng, D., Newgarden, K., & Young, M. F. (2012) Multimodal analysis of language learning in World of Warcraft play: languaging as values-realizing. *ReCALL*, **24**(3): 339–360.