

CHALLENGES IN TRANSCRIBING MULTIMODAL DATA: A CASE STUDY

Francesca Helm, Università degli Studi di Padova

Melinda Dooly, Universitat Autònoma de Barcelona

Computer-mediated communication (CMC) once meant principally text-based communication mediated by computers, but rapid technological advances in recent years have heralded an era of multimodal communication with a growing emphasis on audio and video synchronous interaction. As CMC, in all its variants (text chats, video chats, forums, blogs, SMS, etc.), has become normalized practice in personal and professional lives, educational initiatives, particularly language teaching and learning, are following suit. For researchers interested in exploring learner interactions in complex technology-supported learning environments, new challenges inevitably emerge. This article looks at the challenges of transcribing and representing multimodal data (visual, oral, and textual) when engaging in computer-assisted language learning research. When transcribing and representing such data, the choices made depend very much on the specific research questions addressed, hence in this paper we explore these challenges through discussion of a specific case study where the researchers were seeking to explore the emergence of identity through interaction in an online, multimodal situated space. Given the limited amount of literature addressing the transcription of online multimodal communication, it is felt that this article is a timely contribution to researchers interested in exploring interaction in CMC language and intercultural learning environments.

Language(s) Learned in this Study: English

Keywords: Computer-mediated Communication, Research Methods, Virtual Environments, Multiliteracies

APA Citation: Helm, F., & Dooly, M. (2017). Challenges in transcribing multimodal data: A case study. *Language Learning & Technology*, 21(1), 166–185. Retrieved from <http://llt.msu.edu/issues/february2017/helmdooly.pdf>

Received: November 19, 2015; **Accepted:** October 4, 2016; **Published:** February 1, 2017

Copyright: © Francesca Helm & Melinda Dooly

INTRODUCTION

Identity has become an important field of research in applied linguistics and second language acquisition. As far back as 1995, Norton Peirce called for the development of a “concept of the language learner as having a complex social identity that must be understood with reference to large and frequently inequitable social structures which are reproduced in day-to-day social interactions” (p. 579). In 2011, a state-of-the-art article in the journal *Language Teaching* acknowledged the establishment and rapid growth of identity as an area of inquiry (Norton & Toohey, 2011), witnessed by the fact that identity had also begun to appear in most handbooks and encyclopedias of language learning and teaching (e.g., McKinney & Norton, 2008; Norton & Toohey, 2002; Ricento, 2005). At the same time, it has also been recognized that most research on identity has been carried out in the field of second (L2) language learning and, in particular, immigrant learners (Norton, 2000/2013; Pavlenko & Blackledge, 2004). Much less work has been carried out in the foreign language classroom (Block, 2007/2014), or in the field of computer-assisted language learning (CALL) although it is recognized that “understanding how learner identities...are constructed and enacted in particular contexts can be useful in designing and implementing technology-mediated language learning in education” (Gee & Lee, 2016, p. 166). Much of

the work which has been carried out regards language learners' identity work in social networks (e.g., Chen, 2013; Klimanova & Dembovskaya, 2013; Schreiber, 2015).

The work of identity scholars such as Bucholtz and Hall (2005) and Zimmerman (1998), who adopt a social linguistic approach to the study of identity as it emerges through interaction, has been recognized as a potentially valid basis for the study of identities in online contexts (Thorne, Sauro, & Smith, 2015). However, these frameworks were developed for the analysis of identity in face-to-face contexts, hence the methodological challenge for CALL researchers lies in adapting these frameworks to explore interactions which take place in online contexts as what we see and do on the phone, tablet, or computer screen is highly mediated, filtered, and designed through the technology itself (Kern, 2014). Interaction in online environments is quite different from face-to-face communication, for in any online medium, temporal and spatial relations are transformed (Malinowski & Kramsch, 2014). Also, the many available online environments and tools differ considerably from one another. It is no longer sufficient to distinguish computer-mediated communication (CMC) from face-to-face communication for it has been widely acknowledged that there is no single form or genre of *Netspeak* or CMC, just as there is no single form or genre of spoken or written communication. Facebook has different affordances and constraints for interaction and identity work from, say, Twitter, Tumblr, Whatsapp, Skype, or Snapchat and these will be reflected in the interactions and identity work taking place (Chen, 2013).

Herring (2007), one of the most influential scholars on computer-mediated discourse (CMD) identifies two main types of influence on CMD: *medium* (i.e., the technological aspect which includes aspects such as synchronicity, message transmission, size of message) and *situation* (i.e., the social context, with factors such as information about participants, their relationships to one another, their purposes for communicating). Kern (2015) also emphasizes the situated, relational, and mediated dimensions of meaning making and identity construction in the study of online interactions.

With these concepts of *situatedness* and *mediation* in mind, we discuss some of the challenges faced when studying interaction in online multimodal contexts while exploring identity in interaction. Following the principles of discourse-centered online ethnography (Androutsopoulos, 2008) we address situatedness and mediation by first offering a thick description of the situated context in which the interactions took place and of the functionalities of the communication medium—along with the three available communication modes: visual, oral and textual. We then briefly describe the methodological approach adopted, which, following a social semiotics approach to multimodality (see Jewitt, Bezemer, & O'Halloran, 2016), begins with analysis of each mode separately. However, the main focus of the article is on the challenges of transcribing and representing the interplay between the three modes in a study that, as mentioned above, was concerned with the emergence of identity through interaction in an online, multimodal environment.

CONTEXT OF DATA COMPILATION

The interaction data is drawn from the [Soliya Connect Program](#), established in 2003 in the aftermath of 9/11 (see [Figure 1](#)). The main aim of the program developers was to bridge the gap between young people in Western societies and predominantly Muslim societies and to have them engage constructively with difference. The program runs twice a year (Spring and Fall Semester) and involves partner universities who integrate the program into their institutional courses in various ways. There are two possible strands, one focusing more on academic or political issues and the other on less academic and more intercultural issues, with different options for cross-cultural and individual assignments.



Figure 1. The Soliya website.

The Soliya Connect Program was not developed as an English language program, but it has been adopted in advanced English language courses in various institutes in Europe and the Middle East. The rationale for this is that it is a form of experiential learning, or what could also be considered content and language integrated learning, in which learners are required to use English as a vehicular language to learn more about international relations and divisive issues that relate to participants' identities (Helm, 2013, 2014). It offers students an opportunity for developing their English language and acquiring knowledge about issues which affect the geopolitical and sociocultural relations between Western and predominantly Muslim societies. Furthermore, the program allows participants to develop new online literacies (Guth & Helm, 2010) empathy and intercultural understanding, as well as critical thinking (Bali, 2014).

One of the authors of this article completed an 8-week online facilitation training in 2010 and was then offered her first co-facilitator assignment with a dialogue group. The weekly, 2-hour audio and video dialogue sessions formed a part of a data corpus, some of which is analysed in this study. Written permission was obtained for the use of the data. The names of the participants have been changed to protect their identities.

The dialogue group whose interaction is the object of this study met in Soliya's main meeting room (see Figure 2) over a period of seven weeks. Within the situated context of the online meeting room, the participants had access to one another and could glimpse into the other participants' local physical surroundings through the small webcam picture. This context (main meeting room) can be described as a space in itself with a culture and norms of behavior of its own which are co-constructed by the participants and facilitators through their interactions, mediated by the affordances and constraints of the technological and pedagogic design.

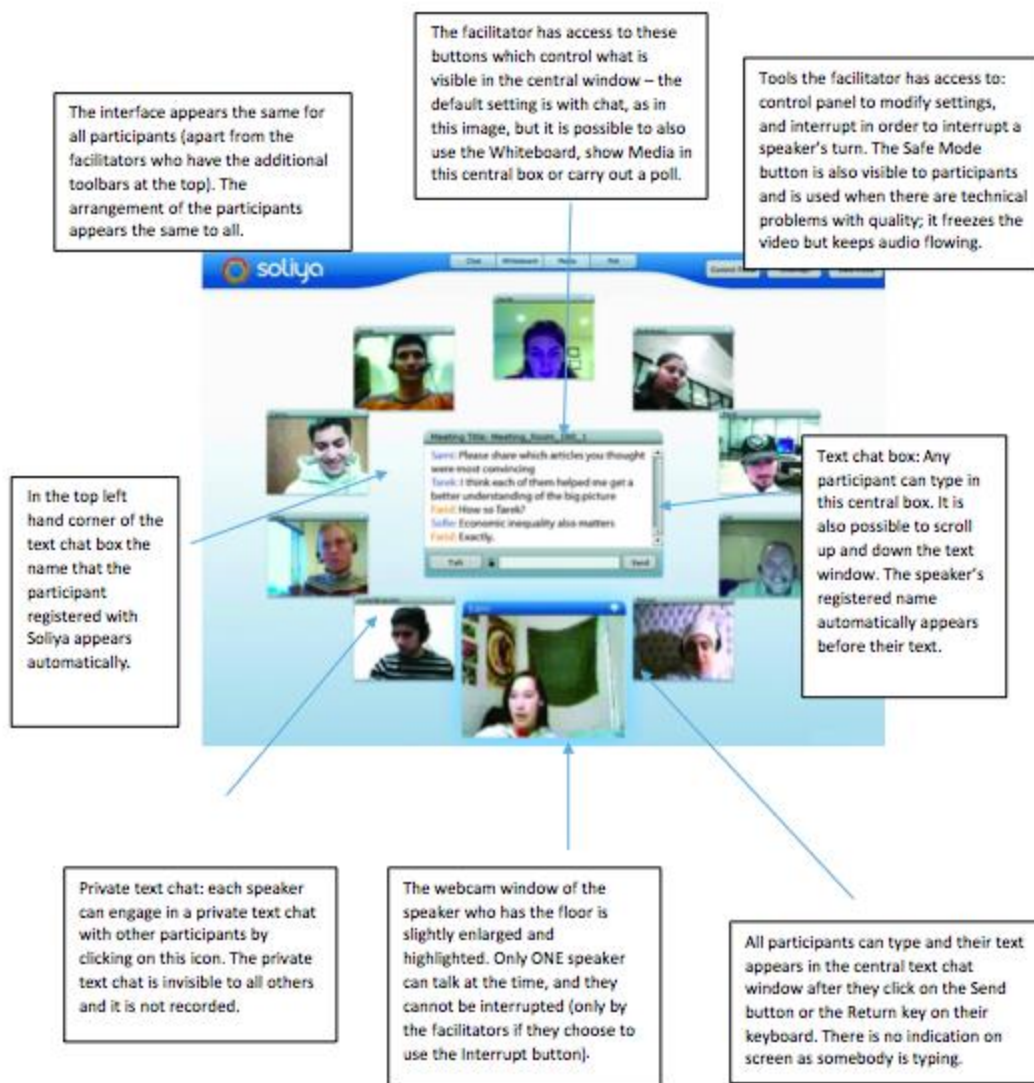


Figure 2. The affordances of the Soliya meeting room.

As can be seen in [Figure 2](#), the situated context of Soliya is an online environment which allows for communication in a range of modes (oral, visual, and textual). This, combined with the fact that there can be up to twelve participants and thus twelve webcam images in the dialogue sessions, makes it a highly complex context for communication where participants have to negotiate the technology as well as language and meaning. As Messina Dahlberg and Bagga-Gupta (2014) observe, these kinds of spaces afford several conversational floors at the same time, and these “are neither exclusive nor additive, but rather multiplicative since they build upon one another” (p. 6). This additional layer of complexity that the mediation of technology adds to the communication inevitably has repercussions for the transcribers and researchers, not least of which is deciding where to focus the analysis.

METHODOLOGICAL APPROACH

As the researchers come from a Conversation Analysis (CA) background (Dooly, 2009; Dooly & Sadler, 2016), CA was intended to be the initial approach for analyzing the data. However, it soon emerged that there were multimodal features of the interaction—due to the mediating effect of the technologies—that

could not be fully accounted for in a CA approach, for example turn-taking mechanisms and sequentiality across modes (Gibson, 2014), not to mention the length and meaning of silence (seconds and even minutes as opposed to fractions of seconds). The approach that we subsequently adopted draws from what Androutsopoulos (2008) calls *discourse-centered online ethnography*, which has been adopted for the study of how identities are constructed through interaction in CMC (Androutsopoulos, 2007; Georgalou, 2016), as well as Herring's (2007) work on CMD and research in the field of multimodality in social semiotics. We also bring into play the outline for approaching multimodal data based on social linguistic traditions by Bezemer and Jewitt (2010). In a social semiotics approach modes are defined as organized sets of semiotic resources used for meaning making. The more a set of resources has been used within a community, the more fully and finely articulated it becomes as a commonly recognizable meaning-making resource (Bezemer & Jewitt, 2010; Guichon & Cohen, 2016).

We first of all privileged *mode* as an organizing principle of identity representation in exploring the affordances and constraints of the media available in this online context for identity construction. The term *mode* is used to refer to many different aspects of communication, for example language is considered a mode—as are speech and writing. There are also visual modes of communication, such as gaze and gesture (Bezemer & Jewitt, 2010; Kress, 2009). The key characteristic of mode within a theory of multimodality is that it is a culturally recognized communication channel or set of semiotic resources. In this case study the modes analyzed are textual (through synchronous, multi-participant, one-way text chat), oral (through the audio), and visual (through the webcams). We are aware of the multiple facets (Herring, 2007) that characterize these broad communication modes, as highlighted in our definition of the platform above. Subsequently, as we explored emerging identities through the interaction itself we adopted an approach which remained multimodal, but was more in line with multimodal research that follows social linguistic traditions (Bezemer & Jewitt, 2010), as the units of analysis were individual turns of spoken interaction and interaction sequences. As will be explained later in the article, we prioritized the oral mode as this was where most interaction was taking place (due to the design of both the online space and the educational program itself) and turned our attention to the text and visual when these were made relevant in the spoken interaction. This is clearly not to say that this approach should be adopted in all studies, but that it seemed the most relevant given the research questions and the situated context of the study.

EXPLORING IDENTITY THROUGH DIFFERENT MODES

Textual Mode

The written text mode was an initial way in to the data, for a record of the written interactions was downloaded at the end of each session. The text chat appeared in the central box of the interface, as seen in [Figure 2](#), and was indeed central to the interactions, largely because the facilitators transcribed a summary of the interaction as it unfolded. This transcribed summary was indexed by the facilitators initially writing the name of the participant followed by a colon, as in [Excerpt 1](#), where Jessica transcribes Brendan's words, and then at the end checks her transcription with Brendan. Indeed, *facilitator-transcriber* was one of the key identities that emerged through the text chat and was certainly the most dominant with over 80% of the text chat seemingly used for transcription.

[Excerpt 1. Using Text Chat for Transcription](#)

Jessica: Brendan: I think we all know or feel that 9/11 def changed the relationship

Jessica: Brendan: wants to ask people in Middle East if there are any events

Jessica: of the same caliber

Jessica: which affected your view of the US

Jessica: right?

Brendan: you got it :)

As for the other participants, some used the text chat frequently (though nowhere near as much as the facilitators), and others barely used it at all. Those who used it more frequently included text chat conventions, such as a wide range of emoticons, abbreviated forms, and repetition of letters, which serve as identity markers (Kern, 2015) and index social presence and the ability to adapt the medium to their communicative needs (Herring 2001, 2015). [Excerpt 2](#) shows how the phatic use of the text chat indexes the group identity and cohesion which emerged through the seven weeks of dialogue as they mark their emotion at the ending of the program.

[Excerpt 2](#). *Session 7, Phatic Use of Text Chat*

Kate: To everyone!! THANK YOU VERYMUCH!

Fadela: luv u alllllllllll

Kate: it was a pleasure to meet you

Alef: I LOVE U GUYS :) bless u all!

Mohammed: thanks guys

Thamena: I want to ask everyone to keep in touch plzzz

Thamena: and I lovee U alll

Doja: it was pleasure meeting you

Jessica: byee Fadela

Denise: bye Fadela!!

Jessica: will definitely kep in touch

Fadela: byeeeeeeee

Jack: bye Fadela!

Fadela: :'(

Alef: bbbye ;)

The challenges of analyzing the text chat mode alone were that sequentiality, which is disrupted even when text chat is the only mode of communication (Anderson, Beard, & Walther, 2010; Gibson, 2014; Tudini, 2010), was particularly difficult to establish since there were no timestamps. Furthermore, without also accessing the audio data it was not clear whether text chat turns were responses to previous lines of text chat or to what was being said through the oral mode. Whilst it was certainly useful as a way into the

data and in defining certain identities of facilitators and participants, analysis of the text chat alone highlighted its very partial nature in this context and the need for multimodal analysis.

Visual Mode: Gaze, Positioning, and Investment

Scholars have begun to explore the affordances and constraints of webcams and have looked at several aspects of the visual mode such as gaze and facial expression (Satar, 2013), framing, positioning, and gesture (Guichon & Cohen, 2016; Guichon & Wigham, 2016; Telles, 2009). Studies have shown the relevance of participants' positioning with respect to the webcam which is constrained by technical features (e.g., built-in or mobile webcams) but which is also often strategically used by participants in projecting a certain image of themselves. Scholars have also analyzed aspects such as proximity to the webcam and framing, which have been found to index interactants' level of trust and confidence with the medium and their interlocutors (Kern, 2015). In [Figure 3](#), for example, we see how some participants have positioned themselves at the bottom-left corner of the window, away from the direct gaze of other participants. In terms of identity work the visual mode also offers access to visual identity markers such as gender, clothes, accessories such as caps or headscarves, and background context (public or private space, décor).



Figure 3. Strategic positioning of participants with respect to the webcam.

Challenges in researching the visual in this study regarded the size and quality of the image, which was small and often of very poor quality due to the inequalities in Internet connections and, above all, the number of webcam images to observe (up to 12 at a time). Most studies which have explored aspects of the visual such as gesture and gaze in video communication have involved dyadic communication (Guichon & Wigham, 2016; Satar, 2013; Telles, 2009), where much more emphasis can be given to the visual mode in the communication itself.

Oral Mode

The main mode of communication in this context, however, was the oral mode as the program was

designed around spoken interactions which were led and supported by the facilitators. This mode can deliver paralinguistic cues such as tone of voice, intonation, speech rate, and identity markers such as accent and hesitation. Challenges in transcription—and also in the interaction itself—are linked to the quality of the audio which can hinder understanding. Often this depends on the quality of the Internet infrastructure which, in turn, can depend on geopolitical as well as technical factors.

In order to research the interactions, it was necessary to transcribe the oral data and to prepare a transcript which included not only the spoken interaction but also relevant information about what was happening through the visual and also the textual modes. We discuss how this was dealt with in the next section.

PREPARING A WORKING TRANSCRIPT OF MULTIPLE MODALITIES FOR THE EXPLORATION OF IDENTITY IN ONLINE INTERACTION

A study of the processes of transcription is key to a greater understanding of the research process. Transcription offers a way in to the analysis, and the needs of the researcher can vary during the different phases of the research process, hence justifying the need for different transcripts. Baldry and Thibault (2006) write the following:

Transcription is itself a form of analysis: it is a textual record of the attempts we make to systematize and unpack the co-deployment of the semiotic resources and their unfolding in time as the text develops. Transcription also prepares the way for other forms of analysis. (p. xvi)

A working transcript of the multimodal data, which brought together the visual, oral, and textual modes, was created in order to offer further insights into the interactions taking place and allow for the selection of interactions for more in-depth analysis. The process for creating the working transcript involved several steps. First, the recordings of the audio and video interaction were viewed several times and were subdivided into more manageable sections (categorized by activity or topics that emerged) for transcription and analysis. For each section, a rough transcription was made of the oral interaction taking place with notes on gestures and other aspects of the video screens which particularly stuck out as aspects worthy of further analysis or which were made relevant through the spoken interaction by the participants (the emic approach).

Each video was viewed again to add a timestamp to each line of the text chat transcript so it could be aligned with the audio and video transcript. This became the multimodal working transcript. This working transcript served for analytic purposes, rather than to be published and shared, and was in constant evolution as the research evolved and challenges were addressed.

Creating working transcripts of the sessions presented several challenges along the way, some of which have been addressed by other researchers (above all in the field of CA), although rarely in an online environment as described in this study. As ten Have (2002) puts it, the process of transcribing must be considered “deeply problematic” (p. 1). Some of the key issues which have been identified by researchers are: (a) recording, (b) selection, (c) representations of time and space, and (d) the format of the transcripts and legibility. We have added (e) the definition of turns and representation of silence. Each of these issues are described separately below along with how they were dealt with in the case study we have described.

Recording

One of the problems ten Have (2002) identifies in transcription is that the processes of recording then transcribing interaction are necessarily selective; they do not provide a full record of what went on. As Ochs (1979, p. 44) points out, “transcription is a selective process reflecting theoretical goals and definitions”. The same can be said of the way in which technology is used for capturing the data: it is selective by nature. In this particular study, the recordings of the audio and video were made by the Soliya organization for internal monitoring, quality control, and research purposes (the participants were fully

informed and gave their consent). The recording captured the screen of the Soliya interface¹ as the participants experienced it during the dialogue and was a relatively faithful reproduction of all participants' viewpoint of the situated context.

There were, however, limitations in this recording set-up. The recording used in this study contained only interactions on the main Soliya interface. There was no record of the private text chats participants may have engaged in with other participants during the sessions or of text chat they may have started writing and then abandoned (Smith, 2008). The recordings also did not include the other windows which participants may have had open on their computers and could have been accessing whilst they were connected to their computers during the session. The researchers also did not have insight into what was happening in the spaces surrounding each participant, the *hors-champ* (Guichon & Wigham, 2016).

For language teachers interested in knowing more about the different language resources used during the interaction, gaining access to the learners' full engagement with the computer interface can be of interest. In such cases, using screen captures of individual computers can provide fuller information of how the learners are engaged with different technology resources. For instance, Video Screen Capture has been adapted for research on writing processes (see Hamel & Séror, 2016) and some researchers have looked at learner strategies such as self-repair in CMC interactions (see Smith, 2008). Depending on the aims of the research study, having full screen capture may or may not be important. The purpose of the current research study was not to explore what participants were doing at their computers whilst taking part in Soliya, but rather to analyze emerging identities in the group interaction, thus recording only the interaction in the platform was not seen as a limitation to the analysis.

Selection

It was clearly impossible to transcribe everything, particularly when we were dealing with video data and when the question of how faithful a reproduction of the interaction a transcript should produce is subject of ongoing debate. When transcribers seek to include too much information in their transcript, it becomes inaccessible even to fellow researchers, let alone the lay reader. This is particularly true when seeking to include non-verbal communication, such as gesture, which is usually represented through symbols in text transcription and requires a legend to explain their use. It is important to acknowledge that there are inevitable losses when rendering spoken and visual interaction through the written language, and inevitably the transcribing process requires selection on behalf of the researcher as to what to include and what to leave out. This must be decided according to the particular interests of the researcher (ten Have, 2002; Ochs, 1979).

In this case study, the predominant mode of interaction was oral, though text chat and visual modes were also used and were integrated into the transcript. The criterion for deciding when to include information about the visual mode in the transcript was based on the interactants making it relevant through the spoken or text chat modes (the emic approach). For instance, as can be seen in [Figure 4](#), Jack curls the index and middle fingers of both hands to indicate air quotes. If we look at the transcript ([Figure 5](#)), we see that he makes this gesture when he says the words *Arab and Muslim world*. This gesture echoes the use of scare quotes that we can see in the text chat window in the same image—where in the first line of text, we can see Jessica had typed a question for the participants (*Is there a conflict between the “west” and the “Arab and Muslim world”*) and also in the last line of text chat (see the right-hand column of transcript in [Figure 5](#)) where Fadela used scare quotes around the word *israel*. With this gesture, Jack seems to highlight potentially problematic terminology, as Jessica and Fadela had done in the text chat. This mutual alignment towards the use of scare quotes (in text) and air quotes (through video) to mark possibly contentious terms links to the identity of a group engaging in dialogue about a controversial topic and aware of the problematization of terminology used to frame the debate.



Figure 4. Screenshot of air quotes gesture.

260	02:36,4	Jack	<p>eer well I said yes for mine but I mean I know that (1s) the: erm the west is er the conflict between the west and the arab muslim [2:47 hand gesture – quote marks] world isn't the BIG conflict at stake there's bigger conflicts (2s) between (1s) internal pla internal affairs I guess? so I mean (1s) yeah I know that erm there is some (1s) unsettled issues between 'hh the west [3:05 uses hand gesture to make inverted commas barely visible] and the east but (2s) it's not the biggest conflict that's there</p>	<p>2:41 Fadela: as i said the main reason of "israel" presense is the west((England))</p>
-----	---------	------	---	---

Figure 5. Transcript of multimodal data including scare quotes in text chat.

Representations of Time and Space

Representation of time is a challenging aspect of multimodal transcription, and is determined in part by the affordances both of the technologies used for the interaction and those used in the transcription. In the transcript of the text chat, which can be seen in [Excerpt 1](#) and [Excerpt 2](#) above, there were no timestamps. Unlike many other platforms offering written as well as audio and video interaction (e.g., Skype, where there is a time stamp for each turn of text), the Soliya platform did not have this, rather just the name of the person typing and the text.

As reported above, a timestamp was added to the text transcript in a second phase of the research, and by comparing the versions with and without timestamp, it was possible to see how the dimension of time contributed significantly to possible interpretations of the unfolding of the text chat. In Version A (see [Table 1](#)) there was a seeming coherence between the lines of text chat; with the exception of the first line in the extract, the subsequent lines make up a coherent conversation. However, Version B (with the timestamps) allowed us to detect significant multimodal features that were not visible in Version A of the text chat. For instance, there was an approximate 30-second delay after Alef typed *tomorrow to Tunisia* and the next line, and a 1-minute delay after Alef typed *to meet Tunisian youth and officials* (highlighted in Version B).

Table 1. Text chat with and without timestamp

Version A	Version B
Ranà: : to write 7 aspects of their identity	1:45 Ranà: : to write 7 aspects of their identity
Alef: we have Hilary Clinton coming7	2:03 Alef: we have Hilary Clinton coming7
Alef: tomorrow to Tunisia	2:09 Alef: tomorrow to Tunisia
Alef: to meet Tunisian youth and officials	2:45 Alef: to meet Tunisian youth and officials
Alef: I'll be there in the meeting	3:47 Alef: I'll be there in the meeting
Thamena: Oh really?	4:02 Thamena: Oh really?
Mohammed: the meeting is about what?	4:19 Mohammed: the meeting is about what?
Fadela: do u think that her visit will help?	4:23 Fadela: do u think that her visit will help?
Alef: Muhammed: I heard that no one in Egypt *accepted to meet her?	4:42 Alef: Muhammed: I heard that no one in Egypt accepted to meet her?
Ranà: thats right alef	5:03 Ranà: thats right alef
Ranà: :)	5:10 Ranà: :)
Mohammed: perhaps because its our problem	5:24 Mohammed: perhaps because its our problem
Mohammed: and we want to solve it ourselves	5:39 Mohammed: and we want to solve it ourselves
Mohammed: we need no help	5:56 Mohammed: we need no help
Alef: JACK: what's the purpose of Clinton's visit to Tunisia and Egypt, now, in your opinion?	6:08 Alef: JACK: what's the purpose of Clinton's visit to Tunisia and Egypt, now, in your opinion?

Representation concerns various features captured in transcripts as regards time measurements but also the *spatialization* of the flow of interaction. As said before, there are multiple layers of meaning making which overlap with one another, and it is important to represent them in the transcription. In the transcription of audio data, particularly when analyzed from a CA perspective, overlap of turns is an important feature to be captured, and conventions have been developed to represent this (Jefferson, 1984; Sacks, Schegloff, & Jefferson, 1974). However, turn-taking in online platforms inevitably has some variability when compared to face-to-face conversations, due to the features and constraints of the technology (Gibson, 2014).

For instance, in the computer-mediated interaction data of this study, overlap between speakers was not possible, for the software did not allow interruption or overlap (unlike Skype, for instance). In order to take the floor a speaker had to click on the *talk* button (see Figure 2), and they had the floor until they released the talk button. Overlap of speaker turns was thus not a concern in this analysis simply because it could not take place. However, overlap of other modes was of interest in this research as the interaction could take place on multiple floors and as one mode could support the other. It could also be used to carry on parallel interactions or subvert the other mode (Sauro, 2009). When working with multimodal online data it was thus important to capture the “transmodality,” or the intersections of the different modes as they “sequentially perforate and interpenetrate each other” (Messina Dahlberg & Bagga-Gupta, 2014, p. 6).

The perforation of different modes was present in our study, thus it was important to capture the way in which the different modes were mediated by the participants as potential communication resources. For example, in one of the extracts, a facilitator (Ranà) used the camera feature to index her discussant *orchestrator* identity by instructing the others to respond to her question through the visual channel using “Thumbs up, thumbs down”. She used one mode of communication, the spoken, but called for a response through a different mode. One of the other group members followed her instructions and produced the expected response by raising her thumb, thereby acknowledging her discourse identity of *addressee* of Ranà's instructions.



Figure 6. Screenshot of cartoonized “thumbs up”.



Figure 7. Screenshot of pixelated “thumbs up”.

Of course, the use of screenshot captures to represent spatiality brought forth yet more issues such as how to maintain the anonymity of the study participants. In this study two technical options were explored, *cartoonizing* (Figure 6) and *pixelating* (Figure 7). In both cases, these inevitably resulted in a loss of detail of more minute facial gestures, though more so in the case of pixelating. For sake of brevity, these are not discussed here. For a discussion of different approaches to the problem see Dooly and Hauck (2012, pp. 150–151).

Format of Transcripts

The format of the transcript can, to a certain extent, support the spatial representation of transmodal communication. Various spatialized formats have been experimented within the literature on multimodal data analysis, from the column format (advocated by Ochs, 1979 for particular kinds of data), to the list format and the partition format (Mondada, 2007)—each of which offers a materialized conception of “the temporal unfolding of talk and conduct and of the sequential organization of their details” (p. 812). In more recent research into online multimodal data, some researchers have prioritized modalities used for communication (e.g., text, emoticon, audio, video) over temporality of the talk (see Dooly & Hauck, 2012).

One of the more predominant formats used by transcribers is the list format whereby time unfolds line by

line representing either successive or simultaneous relations (Mondada, 2007). Gibson (2014) uses the list format, prioritizing the audio mode and inserts numbers in brackets for text chat and other action sequences, such as appearance of typing icon. These are then listed below the excerpt. Messina Dahlberg and Bagga-Gupta also adapted the list format in their 2014 study. Their unit of analysis included the interactional turns (common to CA) as well as the synchronous use of a range of language varieties and modalities negotiated by the participants. Moreover, their adapted CA transcription conventions took into account multimodality with the use of screen captures and figures, with the use of different font styles (bold, italics, or quotation marks) to highlight the language varieties in play, and by making available the written language-in-use. Nonetheless, it must be noted that these procedures, while capturing important details for analysis, can render the transcripts highly complex, and not immediately understandable, particularly to the untrained eye.²

Several analysts have adopted the use of the tabular format (Bezemer & Mavers, 2011; Lamy, 2012; Satar, 2013; Sauro, 2004). The tabular format is commonly used with different columns for each of the different communication modes. This format “constructs temporality on a vertical axis and modal separation horizontally. This provides an impression of how the meanings made unfold synchronously and diachronically, and how they map onto each other” (Bezemer & Mavers, 2011, p. 202). However, whichever element is presented in the left-most column of the table may inevitably be interpreted as intentionally fronting that particular feature (e.g., the gestural mode vs. the speech mode) if we follow the reasoning that Western cultures attend first to the left-hand side of the page (see Bezemer & Mavers, 2011; Dooly & Hauck, 2012; Thibault, 2000).

Lamy (2012) uses a tabular format with the identity of the speaker in the first column on the left, followed by turn number, time, audio-transcript, and text chat. Sequentiality can be vertical and horizontal across the columns as a single conversation can occur across modes, parallel conversations can occur in different modes, and multiple conversations can cross the different modes. Satar (2013) also uses a similar tabular format, interspersed with screenshots of the video screen.

For the working transcript in this study, a tabular format was selected following Lamy (2012), with speakers arranged in rows in a table and modes in columns, as in [Figure 8](#). It is important to acknowledge that spatial and temporal representations are not neutral, but heavily influenced by cultural dimensions. In European literacy cultures, as transcription theorists such as Ochs (1979) have pointed out, ideas are encoded from top to bottom and from left to right (see also Thibault, 2000). Western readers prioritize leftness and hence would look to the left for the inception of a statement or entire discourse. We tend to forget however, that this representation is not universal, nor is it neutral, and that it can influence our interpretations of transcripts. For transparency, decisions on how the transcript is formatted can be made explicit. In our study for instance, we made notes for the visual mode in a column and screenshots were used when necessary, but outside of the table.

Definition of Turns and Representation of Silence

Almost immediately the issue of how to define *turn* became apparent when looking at the simultaneous multimodal data. Would the text chat or the spoken interaction define the turns? Since the main mode of communication was the oral mode, it was decided that the oral mode would be represented as the *dominant* mode in the table and that rows would be defined on the basis of speaker turns rather than text chat turns (despite the fact that the text chat was defined as the starting point for the analysis). Accordingly, the audio column appeared to the left of the text chat column and to the right of the speaker's name. The notes on actions occurring on the video screen were to the right of the text chat, as in [Figure 8](#)**Error! Reference source not found..** A single speaker turn could include several turns of text chat or none.

Related to the issue of turns was the representation of silence, which has not been addressed in recent studies on multimodal communication. If we look at CA (Sacks et al., 1974), human turn-taking is viewed

as so precise that speakers tend to avoid gaps, overlaps, and silences. The mediation of technology in online interactions, however, has a major impact on turn-taking mechanisms (Gibson, 2014). The online interactions in the current study included many silences, some of which lasted well over one minute.

Understanding the meaning of silence in the interactions, as well as the impact of silence on the interaction, became an area of interest as the researchers engaged in the process of transcription. As the working transcription developed, the way silences were transcribed gradually became more complex as we found that there were different types of silence and that these needed to be taken into account, at least in the parts which were subject to close analysis. At times there was a *technical silence* at the end of a speaker turn before the speaker released the floor by clicking on the talk button, in which case the silence was included in brackets at the end of the speaker turn (see Turn 209 in Figure 8). At times somebody had the floor but was not aware of this and did not speak, in which case their name was nonetheless inserted in the speaker column as it meant nobody else could take the floor (see Turn 210). Silences between turns were given their own rows, that is the equivalence of turns in the transcription table, as a visual representation of their significance and to allow for further analysis (see Turn 211). In these cases, silence could be an indication of participants' reluctance to speak or even resistance to being cast into certain identity categories, as we will see in the discussion of the extract. However, silence in the audio mode did not necessarily entail silence in the other modes, as text chat could be taking place when there was silence in the oral mode, as in Turn 211.

If we look at Figure 8 **Error! Reference source not found.**, which is a screenshot of the working multimodal transcript of part of Session 1, we see how this transcript can help us acquire a greater understanding of the interaction and the interplay between the various modes as well as the emergence of identities through the interaction. Ranà, as part of her situated identity of *facilitator*, tried to set up a game through the oral mode (Turn 209). Her spoken turn is followed by a long technical silence when Jessica has the floor but is not talking (Turn 210). Ranà then gives further brief instructions through text chat to write seven aspects of their identity, but as in Turn 209, her instructions are incomplete. In the long silence which ensues (almost 3 minutes) in Turn 211, Alef tries to initiate interaction on a new topic through the text chat, aligning to the situated identity of *participant-initiator of interaction*, also making relevant his own identity as a Tunisian youth with an interest in political affairs as he writes that he is going to meet Hillary Clinton.

The unfolding of the text chat was slow. There were long, silent pauses after Alef's statements, suggesting that he was expecting a response from the co-participants each time he typed a line of text, but two minutes passed before there was a reaction from Thamena, who oriented to Alef's use of the text mode, followed by Mohammed and then Fadela. When Alef wrote that he had heard that nobody in Egypt wanted to meet Hillary Clinton, Mohammed and Ranà both indexed their Egyptian identities through the text chat in their responses to Alef, Ranà stepping out of the institutional *facilitator* identity and making relevant her Egyptian identity. Alef then cast Jack into the category of informed American citizen with an understanding of US foreign policy as he emphatically called on Jack (using capital letters) to explain Clinton's motives for going to Tunisia and Egypt. After nearly 20 seconds of silence Jack responded through the oral mode. This long silence before taking the floor and the hesitation, false starts, disclaimers, and humour that marked Jack's turn as he made excuses for his lack of knowledge on recent events could be seen to index his identity as an uninformed American or his reluctance to express an opinion on the issue with a group of people he was not yet familiar with (this was the first of seven online meetings).

Turn	Time	Speaker	Audio	Textchat	Notes on video
209	0:42,0	Ranà	Well now guys we are going to play a small game together? (...)what do you think (4secs) >What about the game< (1.5secs) Hello? 'hh (3secs) hello (6secs)	0:59 Jessica: Hello 1:04 Jessica: I have to type as you can't hear me	0:55 Jessica has pressed the talk button
10	1:08,4	Jessica			Jessica has floor but is not speaking maybe is typing, after about 30s
211	1:38,1	silence		1:45 Ranà: : to write 7 aspects of their identity 2:03 Alef: we have Hilary Clinton coming? 2:09 Alef: tomorrow to Tunisia 2:45 Alef: to meet Tunisian youth and officials 3:47 Alef: I'll be there in the meeting 4:02 Thameña: Oh really? 4:19 Mohammed: the meeting is about what?	Ranà has floor momentarily, then Alef but no sound 2:20 Mohammed disappears then reappears a couple of times. Jessica looks engaged as she looks at the screen At 3:00 Ranà disappears 3:17 Jessica disappears, 3:37 Ranà re-appears 4:15 Fadela momentarily has floor but says nothing
212	4:23,1	Ranà	(2secs) Is anyone (4secs) is anyone having problem writing the seven aspects about himself? Please write in the chatbox done or err yes	4:23 Fadela: do u think that her visit will help?	
213	4:37,2	silence		4:42 Alef: Muhammed: I heard that no one in Egypt accepted to meet her? 5:03 Ranà: thats right alef 5:10 Ranà: :) 5:24 Mohammed: perhaps because its our problem 5:39 Mohammed: and we want to solve it ourselves 5:56 Mohammed: we need no help 6:08 Alef: JACK: what's the purpose of Clinton's visit to Tunisia and Egypt, now, in your opinion?	4:50 Alef momentarily disappears then reappears 5:50 Ranà disappears 5:55 Alef momentarily disappears then reappears 6:05 Ranà reappears
214	6:27,1	Jack	>it's got nothing to do with how I think since nobody is really doing it with (question) it< urm yeah uhh I uh honestly haven't been too up to date with what kind of like the recent things (...) so I mean I'm not exactly (...)sure [...]		

Figure 8. Working multimodal transcript of Session 1 Turns 209–214.

What was deemed relevant in the extract from the video mode (annotated in the right hand column) were the appearance and disappearance of the facilitators and participants and their attempts to take the floor. These disappearances suggest that technical difficulties were occurring at this point of the interaction and may also account for the silences. Transmodality was indicated in the transcript through the use of arrows which indicated how a topic was communicated through different modes by a particular speaker—in the case of Figure 8, Ranà.

The transcript of this short extract from the first session illustrates the complex interplay between the modes of communication and the importance of representing time and silence. The long silences may reflect technical problems, lack of familiarity of the participants with this new online space and the multiple modes of communication, or a lack of familiarity with one another and the conventions for interaction (i.e., their *novice* identities). Furthermore, the unclear instructions from Ranà and her crossing of modes may have fostered uncertainty as to whether participants needed to orient to her instructions or follow Alef's lead. The silences may also suggest deletions (Smith & Sauro, 2009), with participants writing in the text chat but then deleting, though there is no way of knowing—the software did not have *foreshadowing* (Gibson, 2014), or indication that somebody was typing. This online environment was quite different from face-to-face settings and indeed many other online environments, and the rules or conventions for interaction were being established through their interactions, with the support of the more *expert* facilitators. Arguably, without the adaptation of this new format for transcription the significance of these long silences may not have been revealed.

DISCUSSION AND CONCLUSION

By describing the process and discussing the challenges encountered in the transcription of multimodal data in a research case study that adopted discourse-centred online ethnography and social linguistic multimodal analysis as its methodological points of departure, we have sought to raise awareness of the complexity of this endeavour. We have highlighted the need to recognise the subjectivity of the transcription process which can indeed build on the work of other researchers, but which should also acknowledge the partiality of transcription in terms of what is focused on in the different research studies; the affordances that the communication modes offer; and the sociocultural preferences in how we prepare, read, and interpret transcripts. We have taken examples from a wider study that looks at identity as constituted in the multimodal interaction in the described platform. It is clear that the methodological issues discussed concerning the transcription of online data must be carefully considered because the mediation through technology does have an impact on the way in which these identities are constructed (e.g., affordances and constraints). We argue that the relationship between technology and communication, the medium and the message, should be a key issue in the discussion and research of online interactions with respect to language and intercultural learning. In order to carry out a comprehensive analytical approach, some of the challenges described here should first be acknowledged. The aim of this study was not to offer an attempt at prescribing solutions or frameworks for the transcription of complex, multimodal data. In seeking to explain and justify the decisions made in the process of transcribing multimodal data for this case study which focused on one construct in language learning (identity), we hope to highlight how we can try to be reflexive and critical in our research practices, increasing the transparency and accountability of our work and opening it up for discussion with others.

NOTES

1. To watch sample sessions in progress click on the different files available.
2. In a more recent study, Messina Dahlberg and Bagga-Gupta (2016) have adopted the tabular format.

ACKNOWLEDGEMENTS

This study was partially funded by the KONECT (Knowledge for Network-based Education, Cognition & Teaching) project and financed by the Ministerio de Economía y Competitividad: Proyectos I+D del Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia (EDU2013-43932-P); 2014–2018.

ABOUT THE AUTHORS

Dr. Melinda Dooly holds a Serra Húnter fellowship as teacher and researcher at the Education Faculty of the Universitat Autònoma de Barcelona where she teaches English as a Foreign Language Methodology and research methods courses. Her principal research addresses technology-enhanced project-based language learning in teacher preparation. Her current research interest is project-based telecollaborative language learning and very young learners.

E-mail: melindaann.dooly@uab.cat

Francesca Helm is an English language teacher and researcher at the University of Padova in Italy and recently completed a PhD at the Universitat Autònoma de Barcelona under the supervision of Melinda Dooly. Her research interests are computer-mediated collaboration, telecollaboration, intercultural dialogue, critical approaches to technology and learning, internationalization in higher education and English medium instruction.

E-mail: francesca.helm@unipd.it

REFERENCES

- Anderson, J. F., Beard, F. K., & Walther, J. B. (2010). Turn-taking and the local management of conversation in a highly simultaneous computer-mediated communication system. *Language@Internet*, 7, Article 7. Retrieved from <http://www.languageatinternet.org/articles/2010/2804>
- Androutsopoulos, J. (2007). Style online: Doing hip-hop on the German-speaking web. In P. Auer (Ed.), *Style and social identities: Alternative approaches to linguistic heterogeneity* (pp. 279–317). Berlin, Germany: de Gruyter.
- Androutsopoulos, J. (2008). Potentials and limitations of discourse-centred online ethnography. *Language@Internet*, 5, Article 8. Retrieved from <http://www.languageatinternet.org/articles/2008/1610>
- Baldry, A., & Thibault, P. J. (2006/2010). *Multimodal transcription and text analysis*. London, UK: Equinox Publishing.
- Bali, M. (2014). Why doesn't this feel empowering? The challenges of web-based intercultural dialogue. *Teaching in Higher Education*, 19(2), 208–215.
- Bezemer, J., & Jewitt, C. (2010). Multimodal analysis. In L. Litosseliti (Ed.), *Research methods in linguistics* (pp. 180–197). London, UK: Continuum.
- Bezemer, J., & Mavers, B. (2011). Multimodal transcription as academic practice: A social semiotic perspective. *International Journal of Social Research Methodology*, 14(3), 191–206.
- Block, D. (2007/2014). *Second language identities*. London, UK: Continuum.
- Bucholtz, M., & Hall, K. (2005). Identity and interaction: A socio-cultural linguistic approach. *Discourse Studies*, 7(4–5), 585–614.

- Chen, H. I. (2013). Social practices of multilingual writers in identity networking spaces. *Language Learning & Technology*, 17(2), 143–170. Retrieved from <http://llt.msu.edu/issues/june2013/chen.pdf>
- Dooly, M. (2009). *Doing diversity: Teachers' construction of classroom reality*. Bern, Switzerland: Peter Lang.
- Dooly, M., & Hauck, M. (2012). Researching multimodal communicative competence in video and audio telecollaborative encounters. In M. Dooly & R. O'Dowd (Eds.), *Researching online foreign language interaction and exchange: Theories, methods, and challenges* (pp. 135–161). Bern, Switzerland: Peter Lang.
- Dooly, M., & Sadler, R. (2016). Becoming little scientists: Technology-enhanced project-based language learning. *Language Learning & Technology*, 20(1), 54–78. Retrieved from <http://llt.msu.edu/issues/february2016/doolysadler.pdf>
- Gee, E. H., & Lee, Y. N. (2016). From age and gender to identity in technology-mediated language learning. In F. Farr & L. Murray (Eds.), *The Routledge handbook of language learning and technology* (pp. 160–172). Abingdon, UK: Routledge.
- Georgalou, M. (2016). 'I make the rules on my wall': Privacy and identity management practices on Facebook. *Discourse & Communication*, 10(1), 40–64.
- Gibson, W. J. (2014). Sequential order in multimodal discourse: Talk and text in online educational interaction. *Discourse & Communication*, 8(1), 63–83.
- Guichon, N., & Cohen, C. (2016). Multimodality and CALL. In F. Farr & L. Murray (Eds.), *The Routledge handbook of language learning and technology* (pp. 509–521). Abingdon, UK: Routledge.
- Guichon, N., & Wigham, C. R. (2016). A semiotic perspective on webconferencing supported language teaching. *ReCALL*, 28(1), 62–82.
- Guth, S., & Helm, F. (Eds.). (2010). *Telecollaboration 2.0. Language, literacies, and intercultural learning in the 21st century*. Bern, Switzerland: Peter Lang.
- Hamel, M.-J., & S  ror, J. (2016). Video screen capture to document and scaffold the L2 writing process. In C. Caws & M.-J. Hamel (Eds.), *Language-learner computer interactions: Theory, methodology, and applications* (pp. 137–162). Amsterdam, Netherlands: John Benjamins.
- Helm, F. (2013). A dialogic model for telecollaboration. *Bellaterra Journal of Teaching & Learning Language & Literature*, 6(2), 28–48. Retrieved from <http://revistes.uab.cat/jtl3/article/view/522/571>
- Helm, F. (2014). Developing digital literacies through virtual exchange. *eLearning Papers*, 38, 1–10. Retrieved from <http://www.openeducationeuropa.eu/en/article/Developing-digital-literacies-through-virtual-exchange>
- Herring, S. (2001). Computer-mediated discourse. In D. Tannen, D. Schiffrin, & H. Hamilton (Eds.), *Handbook of discourse analysis* (pp. 612–634). Oxford, UK: Blackwell.
- Herring, S. (2007). A faceted classification scheme for computer-mediated discourse. *Language@Internet*, 4, Article 1. Retrieved from <http://www.languageatinternet.org/articles/2007/761>
- Herring, S. (2015). New frontiers in interactive multimodal communication. In A. Georgapoulou & T. Spilloti (Eds.), *The Routledge handbook of language and digital communication* (pp. 398–402). London, UK: Routledge.
- Jefferson, G. (1984). Transcription notation. In J. Atkinson & J. Heritage (Eds.), *Structures of social interaction* (pp. 191–222). New York, NY: Cambridge University Press.
- Jewitt, C., Bezemer, J., & O'Halloran, K. (2016). *Introducing multimodality*. Abingdon, UK: Routledge.

- Kern, R. (2014). Technology as pharmakon: The promise and perils of the internet for foreign language education. *Modern Language Journal*, 98(1), 340–357.
- Kern, R. (2015). *Language, Literacy, and Technology*. Cambridge, UK: Cambridge University Press.
- Klimanova, L., & Dembovskaya, S. (2013). L2 identity, discourse, and social networking in Russian. *Language Learning & Technology*, 17(1), 69–88. Retrieved from <http://lt.msu.edu/issues/february2013/klimanovadembovskaya.pdf>
- Kress, G. (2009). *Multimodality: A social semiotic approach to communication*. London, UK: Routledge.
- Lamy, M.-N. (2012). Personal learning environments: Concept or technology? Click if you want to speak: Reframing CA for research into multimodal conversations in online learning. *International Journal of Virtual and Personal Learning Environments*, 3(1), 1–18.
- Malinowski, D., & Kramsch, C. (2014). The ambiguous world of heteroglossic computer-mediated language learning. In A. Blackledge & A. Creese (Eds.), *Heteroglossia as practice and pedagogy* (pp. 155–178). Berlin, Germany: Springer.
- McKinney, C., & Norton, B. (2008). Identity in language and literacy education. In B. Spolsky & F. Hult (Eds.), *The handbook of educational linguistics* (pp. 192–205). Malden, MA: Blackwell.
- Messina Dahlberg, G., & Bagga-Gupta, S. (2014). Understanding glocal learning spaces: An empirical study of languaging and transmigrant positions in the virtual classroom. *Learning, Media, and Technology*, 39(4), 468–487.
- Messina Dahlberg, G., & Bagga-Gupta, S. (2016). Mapping languaging in digital spaces: Literacy practices at borderlands. *Language Learning & Technology*, 20(3), 80–106. Retrieved from <http://lt.msu.edu/issues/october2016/messinadahlbergbagga-gupta.pdf>
- Mondada, L. (2007). Commentary: Transcript variations and the indexicality of transcribing practices. *Discourse Studies*, 9(6), 801–821.
- Norton, B. (2000/2013). *Identity and language learning: Gender, ethnicity and educational change*. Harlow, UK: Longman.
- Norton, B., & Toohey, K. (2002). Identity and language learning. In R. B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (pp. 115–123). New York, NY: Oxford University Press.
- Norton, B., & Toohey, K. (2011). Identity, language learning, and social change. *Language Teaching*, 44(4), 412–446.
- Norton Peirce, B. (1995). Social identity, investment, and language learning. *TESOL Quarterly*, 29(1), 9–31.
- Ochs, E. (1979). Transcription as theory. In E. Ochs & B. B. Schiefelin (Eds.), *Developmental pragmatics* (pp. 43–72). New York, NY: Academic Press.
- Pavlenko, A., & Blackledge, A. (Eds.) (2004). *Negotiation of identities in multilingual contexts*. Clevedon, UK: Multilingual Matters.
- Ricento, T. (2005). Considerations of identity in L2 learning. In E. Hinkel (Ed.), *Handbook of research on second language teaching and learning* (pp. 895–911). Mahwah, NJ: Lawrence Erlbaum.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735.
- Satar, H. M. (2013). Multimodal language learner interactions via desktop videoconferencing within a framework of social presence: Gaze. *ReCALL*, 25(1), 122–142.

- Sauro, S. (2004). Cyberdiscursive tug-of-war: Learner repositioning in a multimodal CMC environment. *Working Papers in Educational Linguistics*, 19(2), 55–72.
- Sauro, S. (2009). Strategic use of modality during synchronous CMC. *CALICO Journal*, 27(1), 101–117.
- Schreiber, B. R. (2015). “I am what I am”: Multilingual identity and digital translanguaging. *Language Learning & Technology*, 19(3), 69–87. Retrieved from <http://llt.msu.edu/issues/october2015/schreiber.pdf>
- Smith, B. (2008). Methodological hurdles in capturing SCMC data: The case of missing self-repair. *Language Learning & Technology*, 12(1), 85–103. Retrieved from <http://llt.msu.edu/vol12num1/smith/default.html>
- Smith, B., & Sauro, S. (2009). Interruptions in chat. *Computer Assisted Language Learning*, 22(3), 229–247.
- Telles, J. (2009). Do we really need a webcam? The uses that foreign language students make out of webcam images during Teletandem sessions. *Letras & Letras, Uberlândia*, 25(2), 65–79.
- ten Have, P. (2002). Reflections on transcription. *Cahiers de Praxématique*, 39, 21–43. Retrieved from <http://praxematique.revues.org/pdf/1833>
- Thibault, P. (2000). The multimodal transcription of a television advertisement: Theory and practice. In A. Baldry (Ed.), *Multimodality and multimediality in the distance learning age* (pp. 311–385). Campobasso, Italy: Palladino Editore.
- Thorne, S. L., Sauro, S., & Smith, B. (2015). Technologies, identities, and expressive activity. *Annual Review of Applied Linguistics*, 35, 215–233.
- Tudini, V. (2010). *Online second language acquisition: A conversation analysis of online chat*. London, UK: Continuum.
- Zimmerman, D. H. (1998). Discoursal identities and social identities. In C. Antaki & S. Widdicombe (Eds.), *Identities in talk* (pp. 87–106). London, UK: Sage.