4.

Describing online conversations: insights from a multimodal approach

MARIE-NOËLLE LAMY
ROSIE FLEWITT

1. Introduction

In their Introduction to this volume the editors look at a macro-question which could be summed up as: “What have we learnt since 1987 about conversations and ways to describe them?” In this chapter, by contrast, we address the question at a micro-level: “What does the literature tell us about our particular object of study and how do insights from multimodality and research on computer-mediated conversations help us to structure the way we approach the analysis?”

The object of study in this chapter is a video capture of a technology-mediated multimodal conversation between two native speakers of different languages who have been charged with discussing a set cultural topic and with reporting their conclusions in writing.1 Finding ways to describe and analyse webcam-enhanced one-to-one audio and textual conversations is an interesting aim for teachers of language of culture, as this mode of communication offers a potential arena for structured, semi-structured and unstructured interactions for intercultural learning. But for planned activities to be motivating and successful, they must to some extent conform to the developing style of communicative behaviour that characterises online communication.

With a sharp focus on a 6’41” passage taken from the corpus introduced earlier in this volume, the aim of this chapter is to reflect on the challenges of analysing how the participants use multiple modes of communication in the multimedia orchestration of their online conversation. As a

1. The set topic is presented in the Introduction to this volume. The writing task is spelt out by one of the interlocutors in our extract: Bon, on essaie de résumer [...]. On va le faire par écrit, en fait. V. me demande de le faire par écrit.
technology-mediated social event, the multimodal conversation studied here engages us with issues of mediation and socialisation. As an instructed task, it invites scrutiny of learning processes and outcomes with a view to identifying successful dynamics for online learning dialogues. As a video recorded event, it presents us with issues of data representation. We now address all these aspects as reflected in the literature.

2. Previous research

Mediation research

The Introduction by Stivers and Sidnell (2005) to a special issue of *Semiotica* documents forty years of research on the co-occurrence of talk, gesture and gaze in human interaction. They found that the literature showed “how the vocal modality works,” but that there was an under-representation of research which shows “how the different channels and modalities work together as well as the mechanics that underlie such co-operation” (p. 15). Although the case studies in the special issue itself set out to fill this gap, they include no computer-mediated interactions, and their insights are therefore only of partial help in our task of understanding mediation through video-conferencing.

Mediation² is a feature of all human exchange (Vygotsky 1978; Leontiev 1981; Wertsch 1991), and web-based exchanges fit with the view of human mentality as deeply social (see Tomasello 1999). In computer-mediated contexts, mediational tools include participants, tasks, physical settings, institutional and cultural assumptions, time frames and language (Mercer et al. 2004) as well as technology. Some of these tools represent “new” challenges for analysis, in terms of a distinct difference between online and face-to-face conversations. Firstly, the time frame for interactions is variable: in computer-mediated exchanges, time can be asynchronous, synchronous or “quasi-synchronous” (Garcia and Jacobs 1999). Similarly, the range of media available for communication varies, sometimes restricted to typed text, sometimes accompanied by audio and/or visual exchange, resulting in screens of very different design that offer different potentials for meaning making. The social conditions surrounding elec-

---

². In adopting the term “mediation”, we follow English usage, in which “mediation” refers both to “human-mediated” and “technology-mediated” meaning-making. For a discussion of this duality, see the Introduction to this volume. See also Barbot and Lancien (2003).
Electronic conversations differ from face-to-face exchanges and some users have reported finding online communication easier because of the lack of immediate, visual contact (Becta 2008a). Furthermore, web-based technologies are changing social practices and cultural assumptions about the mediation of meaning (Davies and Merchant 2009).

Despite the potentially transformational characteristics of new technology, Lankshear and Knobel (2008) suggest that computer-mediated text-making practices that are bounded by the constraints of the classroom can all too easily replicate familiar classroom patterns of interaction. Säljö (2004) concurs that “technology has found its way into schools and universities, but its impact has been far less general and revolutionising than the initial enthusiasm led us to believe” (p. 490). However, for Wertsch (2002, p. 106) the introduction of a new cultural tool is necessarily “transformative,” as the change in communicative conditions becomes such that participants cannot simply replicate online what they were doing offline. Additionally, for Kress (2003, p. 32), if cultural resources are to be used successfully for learning, there has to be attention paid “to the materiality of the resources, the material stuff that we use for making meaning” (2003, p. 32, original emphasis). The current authors see desktop videoconferencing as a new cultural tool offering its users plenty of “material stuff” to contend with. Therefore an important focus of our analysis will be to look for signs of such transformation, by studying how participants use “the material stuff” of their encounter to engage in the set task by making use of computer-mediated resources in ways that may be distinct from conventional face-to-face task resolution.

Educational research: multimodal online language exchanges

The conversational event under scrutiny in this chapter occurs in the context of an instructed task, albeit with a simple and informal learning objective: to share insights and write a report. In planned online conversational events set up for educational purposes, the aim is to generate specific learning outcomes. With some exceptions (e.g., Mori and Hayashi 2006, to which we return below), language-learning researchers are interested in effectiveness, and look for it by scrutinising dimensions known to influence learning outcomes, such as group dynamics (Vetter and Chanier 2006; Reffay and Chanier 2003), task design (Rosell-Aguilar 2005; O’Dowd and Ritter 2006; Hampel 2006; Hauck and Youngs 2008), affective variables (Kupersmith 1995; Payne and Whitney 2002; Hauck and Hurd 2005; Develotte 2009) and “electronic literacy” (Hampel et al. 2005; Hauck and Youngs 2008). Insufficient integration of the above-mentioned dimensions can be a block to effectiveness. Insights from these studies are relevant to the
analysis in our chapter, not so much as discrete factors of measurable communicative success but as contributions to the overall psycho-cognitive and technical ecology within which our participants are situated.

**Meaning-making, affect, the computer and the body**

Early research on technology-mediated conversations, of necessity limited to forums and chats, has described the emotional impact that e-multimodality can have on participants in exchanges (Walther 1996). Some studies have identified an inhibitive effect (see Peterson 1997 for an overview of research) whilst others hail a liberating impact (Warschauer 1998). What much of this research has in common is a psycho-social approach centred on the level of comfort or discomfort of the individual. In contrast, only a few examples can be found of a social-semiotic approach to affect in online exchanges. Now that desktop-based technology is providing faster, more varied ways for groups of non co-present conversationalists to communicate, we may hope for more such research to emerge. Examples of issues concerning affect in multimodal settings might include participants’ use of their bodies as a socio-semiotic resource when videoconferencing (or the non-use of their bodies, as in Lee 2007) and the mediation between emotion, bodies and technologies. Wilden’s (2007) study shows how the affective content of a web-telephony conversation was inflected by the material specification of the audio channel opening mechanism, (preventing laughter from being heard by the distant partner) as well as by the subject’s local physical context, as she sat at a computer in a room with others who, though not part of her conversation, on hearing her off-channel laughter, were directly accessing some of its emotional content. New “wearable” technologies (such as headphones, virtual reality applications and “smart” textiles) draw on multiple senses to position users in specific ways as interconnected, embodied beings, and research has found that individuals both engage with, and attempt to insulate themselves from, these new forms of social immersion (Cranny-Francis 2007).

Crucial to our analysis is the notion of “intersubjectivity” – the psychological characteristic of recognizing and taking into account others’ subjectivity. As web users engage in multi-sensory exchanges, they can experience a strong sense of intersubjectivity, which in turn can have an emotional and motivating impact on learning (Becta 2008b).

In keeping with our own objectives, Mori and Hayashi’s (2006) interest in the significance of non-verbal parameters in native/non-native conversations is more semiotic than praxeological, leading them to enquire “how the interlocutors build a common frame of reference which enables them to accomplish understanding when they discover they do not share
common ground” (p. 197). The concept of embodiment in second language exchanges is central to their enquiry but their detailed literature review indicates that little attention has yet been paid to meaning-making via the body’s interaction with the computer, a remarkable exception being the work of Jones (2004) and his focus on the physical environment of the computer-user. Following his line of enquiry, our interest in this chapter is on the interplay of two computers, two bodies, two sets of physical environments and two potentially differing sets of institutional and cultural expectations, since this is the complex backdrop to our current study.

Data representation research

Thibault (2000) was an early voice stressing how the format a researcher chooses for representing multimodal data may exert an influence on the way the data are interpreted (see his critique of tabular formats in Thibault 2000, p. 328). Mondada (2006) discusses three different illustrations of a work meeting: “a) in the form of a transcript based on the linearity of talk-in-interaction [...] b) in the form of transcripts based on a timeline to which talk and action are referred and synchronised and c) in the form of screen shots” (Mondada 2006, p. 119). Again with a focus on face-to-face rather than computer-mediated interaction, Plowman and Stephen (2008) discuss how alternative forms of representing video data can illuminate the variety of ways through which interaction is understood and analysed. They caution that “the use of video necessarily means a focus on the visible and may mean that other, more distal, forms of interaction are overlooked” (p. 562). As Flewitt et al. (2009) are obliged to conclude, after looking at a variety of studies that have used temporal, rhetorical, micro-textual and macro-textual frames, “satisfactory ways [have not] as yet been found to combine the spatial, the visual and the temporal within one system in ways that take account of the perceptual difficulties that inevitably arise when attempting a simultaneous reading (p. 46). As we now turn to the problem of analysing our video extract, the choice of a unit of analysis, dependent as it is on the ways of “reading” that are available to analysts, remains for us a pragmatic decision.

3. Rationale for the choice of data to be analysed

Whilst we consider evidence from across the data set, our analysis focuses on a 6’41” extract (“micro-segment”) of a 25-minute conversation (“macro-segment”) between Andrew (a US native) and Céline (a French native). The reasons that motivated us in selecting this micro-segment are linked
to the way we structured our object of research earlier: a technology-mediated social event (engaging us with issues of online mediation and socialisation), a task (inviting scrutiny of task outcomes), and a multimedia object (presenting us with issues of data representation). These three dimensions shape our approach to the micro-segment.

Socially, the micro-segment is unique within the data because it involves three conversationalists rather than two, as the principal interlocutors (Andrew and Céline) are joined by a third (Jean) who “speaks” to Céline via Windows Live Messenger (MSN) messages. Jean’s presence is made visible in the form of system-generated data (MSN frames and automatic text) and user-generated data (Jean’s typing and the images that he chooses to display). What is also of interest in the social dynamics of the micro-segment is that the trio has unequal access to the exchange and to meaning-making resources: Céline can hear the announcement of a MSN message and can see Jean’s input, whereas Andrew can only hear the tell-tale “bongs” of a new MSN message and see Céline’s facial response to it. He cannot see the message itself or accompanying images. Jean has the most limited access to the exchange, since he sees only Céline’s written message and has no way of knowing what the wider socio-educational context of Céline’s actions is. Figure 1 gives an idea of how participants are represented to each other during this three-way interaction.

From an educational technology perspective, the micro-segment offers a critical angle of study: task-design. The segment covers the moment when
Andrew and Céline end the semi-directed phase of the task, in which they had to hold an informal conversation orally, and they begin the directed phase, in which they had to agree on a written summary of the conversation. In language and intercultural learning, the purpose of consensus-building tasks is to trigger negotiation of meaning. In our data therefore, we might expect to see how the negotiation of meaning takes place, a semiotic phenomenon of particular interest, showing the mediational choices that the participants make in carrying out their task. Specifically, we want to see how meaning-making occurs in the conditions of “multimodal density” that prevail at this point, to borrow Norris’s (2004) phrase for the number of different modes being simultaneously co-orchestrated. We see this as an example of multimodal density because a) Céline and Andrew are using both audio and chat when constructing their joint summary and b) Jean’s MSN messages appear on screen, irrelevant to the task and disruptive of it, and requiring consensual treatment by the two principal interlocutors.

4. Description of the micro-segment

Phase 1 (0′40″): The first phase of Céline and Andrew’s conversation is defined in relation to the task-oriented focus. The phase starts with Céline’s words ok bon on essaie de résumer un peu [ok right let’s try and do some summing up] and ends with her comment je viens de recevoir un message [I’ve just had a message]. In this phase, they each have five turns, in alternation. The content of the exchange is metalinguistic and retrospective in the first six successive turns (e.g., qu’est-ce qu’on a dit on a commencé avec quoi? [what did we say where did we start?]). It is organisational/prospective in the next four turns, relating to what they will report and which medium they will use to do so. An incident – the unexpected appearance on screen of a message seen by Céline and unseen by Andrew – brings the task to a temporary stop.

Phase 2 (0′44″): In this phase, two MSN messages from Jean occupy Céline and Andrew. The task focus of the analysis is lessened but does not disappear, as Céline attempts to deal with the interruption while continuing with the summary (on résume on résume [let’s sum up let’s sum up]). The social and socio-affective focus is brought to the fore, manifested linguistically (un message attend au secours je ne peux pas le fermer [a message wait help I can’t close it]), paralinguistically (smiles and laughter at several points) and technologically (Céline and the researcher, but not Andrew, can read Jean’s two messages, from which it appears that he is a friend of Céline’s). The use of multiple modes for meaning-making is considerable, as Céline multitasks across visual and verbal media, and Andrew supports
her, although they have unequal access to meaning-making resources (six resources for Céline and five for Andrew).

Phase 3 (4′57″): the final phase sees the participants’ return to the collaborative summary-production task. Multimodal meaning-making continues with audio, video and chat, but without the MSN dialogue box and related system sounds.

5. Framework and method for analysis

Our objective is to give an account of conversational meaning-making in this environment, which sees the multimodal co-orchestration of several semiotic systems, accessed simultaneously by each of the three interlocutors through a computer, a headset and a physical room with a computer in it, all of which are “represented” to each other rather than present. To fulfil this objective we shall consider how diverse semiotic systems (including, for example, language, body orientation, gestures and gaze, for which see also Cosnier and Develotte, this volume) index how the participants negotiate meaning within this environment.

A framework that we have found useful to this purpose is Scollon and Scollon’s (2003) concept of geosemiotics, a type of semiotic analysis paying particular attention to where discourses occur within an environment, on the assumption that the location of these discourses is an element of meaning-making with consequences for the overall semiosis of an event. For these authors, semiotic events are environmentally structured into three orders: the interactional order, the visual order, and the place order. The interactional order has to do with how language is positioned in and positions us within the world, and is split into sub-orders such as time, perceptual spaces, interpersonal distance and personal front. The visual order determines the implications of using particular visual objects as part of meaning-making. Finally, the place order determines the semiotic implications of where the social interaction occurs and where signs/indexes are placed within the environment used by the meaning-makers (For a different conceptualisation of the spatial context, see Marcoccia, this volume).

This scheme, to which we have added one refinement concerning sound, has allowed us to account for the complexities of meaning-making in the electronic multimodal environment used by our participants. Auditory resources are included by Scollon and Scollon as part of the perceptual space of the interactional order, yet sound figures saliently in our data, functioning both linguistically and paralinguistically as well as in systems-generated effects (two different kinds of musical alerts). In the multi-media environment...
under study, sound is a recurring characteristic not only of the interactional order, through the presence of the human voice used online as an interactional tool, but also as part of the place order, as discussed later.

Representationally, the micro-segment selected for analysis has been challenging, as it involved creating an account of data from a range of sources: the webcam channel, the audio channel, the chat window, the MSN window, system-generated sounds, and visual data (text and images), both system-generated and user-generated.

Two additional difficulties present themselves in our representational task. The first is the need to convey the difference between what the researcher sees and what each interlocutor sees (for example, Céline sees her own keying in and her own changes in the chat window before she clicks “Send”). The researcher is privy to this input, but Andrew is not: he sees only the “sent” version. The second challenge, well documented in multimodality research offline and online (Thibault 2000, Mondada 2006), is to ensure that the synchronicity and the sequentiality of events are accurately represented. This requirement is challenging not only because the events that we wish to analyse are sometimes of extremely short duration (less than half a second), but also because, to use Scollon and Scollon’s (2003) terms, concomitant events can involve monochronic phenomena (e.g., the vanishing of a frame from the screen) as well as polychronic ones (e.g., the sustained chords of a musical alert).

6. Analysis

We will now use examples from the recorded conversation between Céline and Andrew to demonstrate how the combined frameworks of interactional, visual, place and sound orders can help us to describe conversations that are negotiated through the visual, audio and physical “material stuff” of online exchanges, and through the interaction of human minds and bodies with these technologies.

The interactional order

The interactional order can be divided into sub-orders, as previously mentioned, and here we focus on personal order (how and when interlocutors smile, frown, offer a direct gaze, etc.) and on how language positions interlocutors in computer-mediated exchanges. The semiotic purpose of studying interactional sub-orders is to understand what consequences any variations may have on meaning-making during different kinds of human encounters.
Given that our focus is on conversational encounters in computer-mediated settings, our purpose is to see whether, in these settings, conditions occur that are different from face-to-face encounters, and to find evidence of their impact on meaning-making by the conversationalists.

In two respects, the participants in our micro-segment were operating in different conditions from those in face-to-face conversations. Firstly, they could see their own image as well as their partner’s, and secondly the way that gaze was perceived by the interlocutors was an artefact of the technologies they were using. Our data do not contain primary evidence of the effect of these conditions on meaning-making, but post-recording semi-structured interviews provided us with information on the participants’ feelings about these situations. For both participants, one problem with seeing their own or the other’s image was the unfamiliar direction of the gaze.

The camera’s distorting effect on gaze can be appreciated by studying the sign-on sequence at the beginning of the conversation, when participants had agreed to speak four sentences while gazing at the referenced objects. Figures 2 to 5 show Andrew’s image and gaze direction as it was displayed on both participants’ screens during his sign-on sequence:

Figure 2 – *Je regarde mon interlocuteur*  
[I’m looking at my interlocutor]

Figure 3 – *Je regarde mon image*  
[I’m looking at my image]

Figure 4 – *Je regarde la fenêtre du chat*  
[I’m looking at the chat window]

Figure 5 – *Je regarde la caméra*  
[I’m looking at the camera]
The only image that approximates closely to what a face-to-face conversationalist might perceive as a direct gaze is Figure 5, when Andrew is looking at his webcam and not at Céline’s on-screen image. In contrast, in Figure 2, he is looking at Céline’s image, yet from her viewpoint (which is also the researcher’s, as the images in Figures 2 to 5 are stills from Céline’s screen) he displays what might in face-to-face conversation be interpreted as an evasive gaze, an avoidance of eye-to-eye contact.

In interview, Andrew as well as Céline (as noted also by Cosnier and Develotte) mentioned the affective effects of gaze direction in filmed interaction:

Andrew (on seeing his own image)

*Quand je me suis vu j’ai je me suis rendu compte que le caméra était un peu décalé voilà, je faisais des ajustements.*

[When I saw myself I realized that the camera was a bit offset, so I made some adjustments.]

Céline (on seeing her own image)

*On regarde [sa propre] image mais elle est faussée parce que nous on se voit toujours avec les yeux pas au bon endroit donc ça fait un peu c’est un peu spécial. [...] En plus de me voir, j’entendais ma voix donc c’était vraiment perturbant.*

[... you look at (your own) image but it’s weird because you always see yourself with your eyes not in the right place so that makes it a bit it’s a bit weird. (…) As well as seeing myself, I could hear my voice so it was really disturbing.]

Andrew (on the direction of the gaze of his partner)

*Je pouvais pas le voir en face parce qu’il regardait [...] mon image qui était un peu en bas à droite je pense comme le sien était pour moi, donc c’était un peu bizarre [...] mais ce qui était bizarre c’était que son image n’était pas devant le caméra donc euh il regardait en bas à droite. [...] Donc euh c’était j’imagine qu’il a dû sentir ça aussi mais je ne sais pas.*

[I couldn’t look her in the face because she was looking (…) at my image which was a bit lower on the right I think like hers was for me, so it was a bit weird (…) but what was weird was that her image wasn’t in front of the camera so um she was looking down to the right. (…) So um it was I imagine that she must have felt that as well but I don’t know.]

How interlocutors respond to the challenges of using webcams can in turn be associated with their status as experienced users or as “newbies”: downward gaze signifying humble, shy, inexperienced users and gaze at the camera signifying confidence and experience (Adami 2008). In his interview, Andrew mentioned that although he had previously used typed messages

---

3. We assume that Andrew refers to his partner (Céline) in the masculine (*le, il*) because he construes "partenaire" as a masculine-only noun.
online, this was his first experience of using a webcam and chatting online: *c'était complètement nouveau* [it was completely new], but with time he found it easier to use the electronic tools at his disposal. Céline also reported that she had little experience of webcams — although she was already familiar with MSN and Skype. Gaze direction and the affordances of a webcam used during online interaction therefore appear to require skill and practice on behalf of the users in adapting to both computer software and hardware.

Based on our data, we cannot claim that the technology produces effects that can be shown to play a part in the participants’ meaning-making in the conversational process. Yet we can say that in webcam-enhanced conversations, the semiotic resources relating to “personal front” are managed differently than in face-to-face conversations, and that this difference can cause perplexity ( “bizarre”, “spécial”) and discomfort ( “perturbant”). Furthermore, users’ familiarity and skill with the technical tools at their disposal can have an impact on their status within a user group.

<table>
<thead>
<tr>
<th>CÉL</th>
<th>&amp;ok: bon on essaye de: résumer un peu: enfin() de ce qu’on a dit hein&amp; ok right let’s try to do some summing up I mean () of what we said ok</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOTH</td>
<td>&amp; &amp;</td>
</tr>
<tr>
<td>AND</td>
<td>&amp;euh d’accord&amp; er ok</td>
</tr>
<tr>
<td>AND</td>
<td>((nods))</td>
</tr>
<tr>
<td>CÉL</td>
<td>((C looks at A’s image))</td>
</tr>
<tr>
<td>AND</td>
<td>((A does not look at the screen but looks to his right))</td>
</tr>
<tr>
<td>CÉL</td>
<td>((laughs)) de quoi on a parlé? what did we talk about?</td>
</tr>
<tr>
<td>AND</td>
<td>qu’est-ce qu’on a dit on a commencé avec quoi what did we say where did we start?</td>
</tr>
<tr>
<td>CÉL</td>
<td>((C looks at A’s image))</td>
</tr>
<tr>
<td>AND</td>
<td>((A does not look at the screen but looks to his right))</td>
</tr>
<tr>
<td>CÉL</td>
<td>&amp;euh () on a parlé euh un peu de nos études&amp; we talked a bit about our studies</td>
</tr>
<tr>
<td>AND</td>
<td>on a parlé des différences je pense we talked about the differences I think</td>
</tr>
</tbody>
</table>

Table 1 – A representation of speech, gaze and body movement

With regard to the impact of technology on how the participants mediate action through language use, Bakhtin (1988) proposes that social and historical contexts create and define social systems with their own genres of specific language features. This socio-cultural perspective affirms that social activities cannot be analysed separately from the artefacts that mediate them.
Webcam supported computer-mediated exchanges can therefore be viewed as a new sub-genre of interaction, characterised by a sequence of speaking turns that are often uninterrupted, and are carefully negotiated through the physical cues of gaze and body movements (shown here as conversational turns on a par with linguistic turns, with curved lines denoting simultaneity. For an alternative representation of simultaneity see Appendix, p. 94)\(^4\).

This dialogic perspective on interaction brings intersubjectivity into the analysis. In her interview, Céline explained that although words were used to open a dialogue, the presence of the webcam image helped her to begin to recognise and take into account the other person's subjectivity:

Céline:

\[
l'image \ évite \ beaucoup \ de \ malentendus \ et \ justement \ on \ a \ bien \ compris \ comment \ réagit \ l'autre \ (.) \ on \ arrive \ assez \ rapidement \ à \ son \ humour
\]

[the image avoids a lot of misunderstandings and you understood precisely how the other reacted (.) you get to understand their humour quite quickly]

Both Andrew and Céline found only being able to see the other person's head was semiotically constraining. For Céline, the restricted modalities offered by webcam exchanges “took away the keys to her understanding.” For example, Céline mentioned that if her interlocutor was disturbed by a noise that she could not hear because she was not in the same room, then she would not know why the person was less receptive and might mistakenly think it was her fault. However, she noted that hand gestures could often be seen via the webcam, and this was particularly important in an international context, where prosody and humour are manifested differently, since interpreting hand and facial gestures can help to reduce potential misunderstandings. For Céline, therefore, embodied gesture appeared critical to establishing both intersubjectivity and a common frame of reference.

This raises the notion of presence and absence in online conversational exchanges. Whilst an interlocutor may be represented on-screen as being present (for example, through an icon symbolizing that they are online or through a webcam image), they may also be regarded in the interactional order as absent: physically absent from the experience of the other. For Andrew, the distance between interlocutors and the anonymity of on-screen interaction was a liberating factor: on ne connaît vraiment pas les gens donc on pourrait dire un peu n’importe quoi et ça fait plus (.) je pense beaucoup

---

\(^4\) In this transcript, the conventions are as follows: ( ) denotes a pause of up to 0’2”; (( )) denotes a comment by the transcriber; & at turn start and turn stop indicates that the speaker is looking at her/his partner’s image during the turn; && indicates that the partners are looking at each other’s images.
plus à l’aise [you don’t really know people so you could say whatever you wanted and that makes you (. . .) I think it makes you feel more at ease.] This perspective introduces the theme of multiple subjectivities and playing with one’s identity in online exchanges with unknown others.

The visual order

The reader is invited to briefly return to Figure 1, which will convey the visual complexity of this computer-mediated conversation. Each of the two screens that mediate the exchange is composed of several demarcated sections, with a large text box to the left and extending into the centre. This displays differently coloured system-generated and user-generated messages that include text and emoticons against a white background, with iconic symbols that signify the presence of a message and/or interlocutor. Underneath is a smaller text box where the participants type and send their messages, which offers a selection of emoticons provided ready-made within the software. To the right of the screen are the webcam images of both participants, one above the other. At the base of the screen, we can see that Céline has three different programmes running, including MSN, a conversation with Andrew and a conversation with Jean. There are therefore multiple relationships being conducted simultaneously, all constituting the visual order of each participant’s screen.

There is not space in this chapter to analyse this level of visual complexity, so in our analysis, we turn to the visual order to structure our observations concerning how participants manipulate visual objects to pursue their conversational objectives. For example, in a conversation between two people with a mobile phone placed on the table between them, some action is likely to be taken if the mobile phone screen lights up, signalling an incoming call or text message, with a repercussion of some kind on the conversation. An example from our data is Céline’s struggle with the first of Jean’s messages (beginning of Phase 2). The significance of the “delete” cross in the right hand corner of the MSN message box is familiar to Céline and her first response is to click on this cross, an action followed by the disappearance of the box. This action of Céline’s is technically appropriate. However it is conversationally inappropriate: it has not ensured the disappearance of Jean from the conversation (as is clear from his subsequent repeated returns), nor has it ensured Céline’s continued focus on her assigned partner (as we can hear, from her continued apologies to him). On the necessity for a double closure, both conversational and technological, see also Liddicoat, this volume. From this evidence we conclude that in this environment, it is possible to have a decisive impact on the conversational
insights from a multimodal approach

process (e.g., to allow the conversation to continue, or to end the conversation) through acting on resources in the visual order. In other words, the visual order may determine what goes on in the interactional order.

A further example will show how the visual order is also linked to the temporal sub-order of the interactional order. Our participants have asymmetrical access to visual meaning-making resources, depending on the time-frame \((T_1', T_2')\) under observation. In Table 2, based on data seen from Céline’s screen, we itemize those resources accessible to Andrew (aA) and those accessible to Céline (aC) at time \(T_1'\), which we define as a duration of time when only Céline is acting on visual objects.

<table>
<thead>
<tr>
<th>aA and aC</th>
<th>Chat messages addressed to Andrew after transmission</th>
</tr>
</thead>
<tbody>
<tr>
<td>aA and aC</td>
<td>Information (textual and iconic) about the system, about partner and about self</td>
</tr>
<tr>
<td>aC</td>
<td>Chat messages addressed to Andrew before transmission</td>
</tr>
<tr>
<td>aC</td>
<td>Chat messages addressed to Jean before transmission</td>
</tr>
<tr>
<td>aC</td>
<td>Chat messages addressed to Jean after transmission</td>
</tr>
<tr>
<td>aC</td>
<td>Chat messages from Jean and information (textual and iconic) about Jean</td>
</tr>
</tbody>
</table>

Table 2 – Access to visual resources at time \(T_1'\)

However this analysis of access is only valid for the part of the micro-segment during which Céline has the initiative of writing and handling the visual objects. There is no system-controlled block on Andrew accessing all of these meaning-making resources. Supposing that he took the initiative in \(T_2'\), and assuming we could see his screen, the distribution might look like this:

<table>
<thead>
<tr>
<th>aA and aC</th>
<th>Chat messages addressed to Céline after transmission</th>
</tr>
</thead>
<tbody>
<tr>
<td>aA and aC</td>
<td>Information (textual and iconic) about the system, about partner and about self</td>
</tr>
<tr>
<td>aA</td>
<td>Chat messages addressed to Céline before transmission</td>
</tr>
<tr>
<td>aA</td>
<td>Chat messages addressed to (potential) private correspondents before transmission</td>
</tr>
<tr>
<td>aA</td>
<td>Chat messages addressed to (potential) private correspondents after transmission</td>
</tr>
<tr>
<td>aA</td>
<td>Chat messages from (potential) private correspondents and information (textual and iconic) about these persons</td>
</tr>
</tbody>
</table>

Table 3 – Access to visual resources at time \(T_2'\)

From the comparison of the two tables, it is clear that in this environment, access to meaning-making resources and the conversational processes that
depend upon it, are also dependent on the interrelationship between the visual order and the temporal sub-order of the interactional order.

**The place order**

Finally, we use the place order as a construct that allows analysis of the effects of place and the placing of signs on conversational events. For example, a speaker in a meeting might glance up at a wall clock (a visual sign permanently available to all in the room) or might be made to look at a time warning on a sheet of paper waved by the session chairperson (a sign available only to the chairperson-speaker pair). Each of these time-keeping options has different potential conversational repercussions: for example, in the latter case, the speaker may address the chairperson, acknowledging receipt of the information as well as acceptance of the chairperson’s status, in an “aside” audible to everyone in the room (a set of conditions entailing a richer social interchange than is likely to happen if the speaker merely looks up at a wall clock). In our data, the importance of the place order is evidenced in Céline and Andrew’s deployment of different techniques for upholding their common conversational objective.

Table 4 shows the actions carried out by Céline based on observing the placing of her cursor on the screen during Phase 2. Beneath the Table, Figure 6 illustrates the movements of her cursor at successive times Ta, Tb, Tc etc.

<table>
<thead>
<tr>
<th>Céline’s cursor is positioned...</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ta</td>
<td>on the “delete” cross in the corner of the dialogue box where Jean’s message is displayed</td>
</tr>
<tr>
<td>Tb</td>
<td>at the leftmost point of the chat box (where messages for Andrew can be prepared)</td>
</tr>
<tr>
<td>Tc</td>
<td>over the orange flashing rectangle alerting her that Jean is trying to communicate</td>
</tr>
<tr>
<td>Td</td>
<td>at the leftmost point of the dialogue box where Jean’s messages are displayed, over the “send” button of the dialogue box where she has prepared a message for Jean</td>
</tr>
<tr>
<td>Te</td>
<td>over the “delete” cross in the corner of Jean’s dial box</td>
</tr>
<tr>
<td>Tf</td>
<td>at the leftmost point of the chat box (where messages for Andrew can be prepared)</td>
</tr>
<tr>
<td>Tg</td>
<td>over the “delete” cross in the corner of Jean’s third message</td>
</tr>
<tr>
<td>Th</td>
<td>at the leftmost point of the chat box (where messages for Andrew can be prepared)</td>
</tr>
</tbody>
</table>

Table 4 – Positions of Céline’s cursor during Phase 2
An outline of her cursor movements on the screen during the same period shows that she moves between Box A (Jean’s MSN message box), Box B (the area for chatting with Andrew), Box C (the alert box for MSN, showing in flashing orange when Jean is trying to communicate) and Box D (the dialogue box with Jean), over the entire screen space, in seven separate moves numbered 1 to 7 in Figure 6.

Figure 6 – Movements and positions of Céline’s cursor during Phase 2

Figure 7 – Movements and positions of Céline’s cursor during Phase 3
For comparison purposes, we have looked at Céline’s cursor movements and positions in a subsequent segment of identical duration, in which she has successfully removed the interruptions and she and Andrew are now collaborating in producing their summary. In Figure 7, the area of cursor activity is materialised by a dark ovoid shape, and small white squares represents positions where Céline’s cursor stopped as she typed the agreed summary into her preparation box, occasionally moving backwards and forwards between the squares to correct errors.

Resources in the place order are used by Céline in pursuance of her objective, which is to eliminate disruption and bring to fruition her collaboration on the creation of a written report, through a consensus with Andrew, prepared orally. The juxtaposition of the two figures, showing non-linear deployment of many resources in the place order at a time of conversational difficulty is in contrast with the simpler scheme in Figure 7, when the conversation is uneventfully collaborative, and it materializes the interrelationship between resources in the place order and those in the visual and interactional orders.

A further aspect of place order that emerges as noteworthy in our data is how place, as a physical setting, can index computer-mediated communication. In her interview, Céline reflects on how the physical location of the online exchange – which room it is conducted in and where that room is located within a home or educational institution – may have a significant impact on the nature of what she says and does: when chatting online at home, Céline felt freer than when in school. Céline also referred to physical presence and absence as factors in mediating meaning and understanding others’ perspectives:

Céline:

quand il y a deux personnes dans la pièce cela influera aussi sur le discours (...) par contre on a peut-être plus de mal à comprendre ce que l’autre est en train de vivre (.) on ne sait pas ce qui se passe autour de lui

[when there are two people in the same room, that influences the talk (on-screen) it’s more difficult to understand what the other is experiencing (.) you don’t know what’s going on around him]

The data suggest therefore that the physical location of an online conversationalist in a social world, whether at home, in a public space or educational environment, can index what and how meanings are expressed and
understood, again reinforcing the interdependent relationship between the place and interactional orders.

The sound order

Finally, we move on to consider the sound order, with a focus on the sound track present in the data micro-segment selected for analysis in this chapter. In interview, both Andrew and Céline mentioned the disruptive-ness of the technical difficulties they experienced initially with echo in the sound, and Céline found hearing her own voice played back via the computer rather perturbing.

During Phase 2, in addition to the electronically transmitted sound of the participants’ voices, five system-generated sounds are heard: the triple bong announcement of a MSN message, repeated approximately six seconds later; then a more urgent and longer series of bongs as Jean’s MSN message appears on-screen; then a third triple bong MSN announcement and finally an unidentified system-generated sound is heard by both students, which Céline mentions: *J’entends un bruit* [I hear a noise] and Andrew responds to by pulling a face and adjusting his microphone.

Each of these sounds cuts across the flow of the conversation, and has some consequence for the comfort of the interlocutors. Céline sees the MSN message appear on her screen, hears the accompanying sounds and can see Andrew’s reaction to the sounds. By contrast, Andrew can only hear the message alert and see Céline’s responses to the interruption. On hearing the alert, Andrew leans forward towards the screen, as if to bring his body closer to the sound source, although he is wearing a headset and all sounds reach him through this device, regardless of how close his ear is to the computer. This use of his body (bringing himself closer to the screen and therefore “virtually” closer to Céline, whilst also smiling) also indexes his sympathy for her social position of trying to conduct an educational task whilst being interrupted by a friend. He then reinforces this reassurance through the mode of language: *ah là il y a un autre* [oh here’s another one].

Céline’s response to the message carries a sense of urgency as she orchestrates diverse modes to handle the situation. Table 5 represents her actions (shown here as conversational turns on a par with linguistic turns, with curved lines denoting simultaneity). For an alternative representation of simultaneity see Appendix (p. 94).6

6. The conventions for this passage are: ■■■■ denotes 4 “bongs”; ☼ denotes cursor position over the “delete” cross of the WLM box; □ denotes the disappearance of the WLM box; (0.3) and (0.4) denote pauses of 3 and 4 seconds.
As Céline gives this verbal account of her actions to Andrew, she turns her gaze away from the webcam, alternating her gaze direction between her screen and keyboard as she responds to the message. Here, her body orientation and gaze direction clearly index the temporary backgrounding of her relationship with Andrew.

In the transcript, we have chosen to represent sound through icons that are explained in a Key, and are accompanied by verbal accounts of the nature of the sounds heard on the data sound track. Particular problems confront researchers in the representation of sound, as Ong (1982) points out: “There is no way to stop sound and have sound. I can stop a moving picture and hold one frame fixed on the screen. If I stop the movement of sound, I have only silence” (p. 32).

In this tiny segment of data, we can see how Céline is a multimodal meaning-maker, orchestrating resources from the different orders and sub-orders. This opportunity for multimodal functioning is granted to her for three reasons: a) the technological affordances of the environment allow it; b) there is a chance event in her social environment (her friend Jean happens to be trying to make contact); c) she takes the lead in using the tools, as we know both from her statements in interviews, and from Andrew’s comment to her that he is not used to the French AZERTY keyboard and would prefer her to do the typing. Although Andrew’s opportunities for multimodal meaning-making are limited to condition (a), he too is co-orchestrating multiple tools. He has fewer visual resources at his disposal, as we saw from Table 1, but he makes use of the systems-generated sounds, body movement, gaze and language to assure Céline of his understanding of her difficulties. The two interlocutors are familiar
with these kinds of online interruption, and appear to view them as an acceptable feature of online communication. This begs the question “does this constitute evidence of a shift in beliefs and practices about the learning environment as a place where distant personal and educational relationships overlap?”

7. Discussion

Adopting a multimodal approach to data representation and analysis, we have considered how interactional, visual, place and sound orders index the mediation of meaning during a micro-segment of webcam-enhanced, online conversation. We have demonstrated:

– How the nature and conditions of computer-mediated educational conversations can be very different from face-to-face encounters, and can be viewed as an emerging sub-genre of interaction that requires new sets of analytic descriptors to understand more fully how communication and learning are played out in these environments. The participants in the data studied here were not experiencing the same conversational conditions, as they were sitting at differently resourced computers, with different access to software systems, in different rooms. The free-flow of conversation was highly dependent upon participants’ technical skills, know-how, ease and ability to co-orchestrate diverse semiotic systems. Multiple, overlapping semiotic systems constitute a semiotic aggregate in the mediation of meanings online, where multiple exchanges can be conducted simultaneously with interlocutors in distant and dispersed locations.

– How both embodied cues (e.g., nodding, but also cursor moves) and verbal cues play key roles in the mediation of meaning a) to communicate interactional intent b) to foster intersubjectivity c) to send relational messages, and d) to structure and maintain conversational exchange.

– How the multiple semiotic systems available to conversationalists in computer-mediated exchanges can operate as systems of social positioning and generate power relationships, particularly if the participants have unequal access to and experience of computer hardware and software. In the data studied, we have seen how a particular keyboard layout led to the French native speaker being the “scribe” for the set task.

– That the multiple media that characterize on-screen exchanges can be so diverse and rich that users may find it difficult to manage what is to be attended to. In the case of our data this occurs when the domains of personal life (the MSN message received by Céline) and educational exchange co-occur.
That there is a tight interrelationship, both semiotic and chronological, between the interactional, visual, place and sound orders in the online mediation of meanings. Thus, we see a useful research agenda emerging: to test out, with a large volume of conversational interaction data collected from multimodal environments, the methodological assumption that such data could be best analysed through the synergistic use of conversation analysis, social semiotics and geosemiotics. However, it is a complex task and we argue that there is a need to approach it not with a grand vision, but in a practical manner and one step at a time, “thinking,” as Scollon and Scollon (2003) advise, “in terms of small systems of meaning interacting with each other” (p. 157).

For researchers, our analysis indicates that before settling on a method for representing the multimedia and multimodal complexity of online interaction, they should reflect on the theoretical goals underpinning the research. All choices made with regard to data representation have implications for what can be understood, and as has been documented before (Thibault 2000), no choices are innocent. The choices that we have made in this chapter are practical, and each one was determined by the particular point being represented. However in Appendix (p. 94) we show how alternative choices may inflect the analysis.

Researchers also need to develop technical skills to exploit the potential of data representation offered by computer software. Sadly, though, researchers often remain constrained by current conventions of publishing in print.

From an educational perspective, if we can analyse in detail how learners mediate meaning when they are using the multiple modes and media available to them when co-negotiating tasks in online conversation, then we will be better able to provide educators and software designers with appropriate models of how learners may respond to the technical tools they have available to them. Theoretically, educational tasks can then exploit the potential transformation offered by these multimedia environments, and they can be tailored to the style of conversational behaviour that characterises online communication. However, all learners make self-regulated choices about how they engage with the tools available to them, so in designing instruction educators should conceptualize multimedia features as affordances, as invitations for learners’ engagement, rather than assuming that all learners will engage with them in similar ways.

Deeper understandings of learner perspectives and online performances could be gained through multimodal, ethnographic studies that explore Bourdieu’s (1977) notion of “habitus” in relation to online practices, giving insights into how social beings bring certain dispositions with them that are acted out through multiple modes in their online perfor-
mances. Longitudinal studies in this field are also needed. Indeed Mercer (2008) observes that:

[A]s learning is a process that happens over time, and learning is mediated through dialogue, we need to study dialogue over time to understand how learning happens and why certain learning outcomes result. (p. 5)

We would add, however, that particularly in the context of computer-mediated exchanges, rather than focusing purely on language use, a multimodal approach affords more telling insights into how conversational meanings are mediated through multiple media and diverse modes.
## Appendix

<table>
<thead>
<tr>
<th>Audio channel (Céline’s voice)</th>
<th>Expressive phonation from Céline (C) and Andrew (A)</th>
<th>Body movements from Andrew</th>
<th>Audio channel (system)</th>
<th>Gaze (&amp; -&gt; denotes direction of gaze)</th>
<th>Traces of screen actions (non-linguistic)</th>
<th>Traces of screen actions (linguistic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>attends (0.3)</td>
<td>leans towards screen</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T+3”</td>
<td>C: small chuckle</td>
<td></td>
<td></td>
<td>– A &amp; -&gt; C</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– C &amp; -&gt; keyboard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T+6”</td>
<td>(0.4) bon j’ai envoyé un message alors donc on résume</td>
<td></td>
<td></td>
<td>– A &amp; -&gt; C</td>
<td></td>
<td>WLM message:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– C &amp; -&gt; WLM box</td>
<td></td>
<td>celine-en-hiver says</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– A &amp; -&gt; off-screen right</td>
<td></td>
<td>attends je bosse</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>☽ C’s cursor is above the clickable X in the top corner of the dialogue box.</td>
<td></td>
<td>[hang on, I’m working]</td>
</tr>
<tr>
<td>T+10”</td>
<td>et puis tu</td>
<td></td>
<td></td>
<td>– C &amp; -&gt; chat box</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This transcript is a matrix showing the simultaneity of language, gaze, movement and actions through horizontal juxtaposition (using the material in Table 5, p.90). Contrasting with the choices made in the chapter of using vertical sequencing to represent turns, this matrix allows for greater chronological accuracy, but encourages a reading which foregrounds information in the leftmost column (here the language in the audio channel) to the detriment of information on the right (here the non-linguistic conversational material deployed). Also, choosing a matrix to represent events that are not co-terminous means selecting a time (T) as an “anchor” and arbitrarily deciding that a particular input will be set at T (here, Céline’s first word “attends”), rather than some other (for example the start of the first of the four “bongs” ■■■■, or the moment when Andrew starts leaning towards the screen).