

## **Multimodality in Second Language Conversations Online: Looking for a methodology**

**Marie-Noëlle Lamy**

**The Open University,  
United Kingdom**

### **1. Introduction**

This paper analyses interactions undertaken in technology-enhanced environments by foreign-language learners to enhance their fluency in oral discourse and exercise distance teaching skills. The paper which focuses on such phenomena as turn-taking and face-saving takes Conversational Analysis (CA) as its theoretical starting point but adopts a multimodal perspective that appears to shed greater light on the meaning-making processes involved in interactive digital environments.

The approach is corpus-based and uses examples of conversational moves across modalities taken from a learner corpus (university students of English as a second language). In this respect, the availability since the late 90s of “Internet-telephony” environments offering synchronous voice with near-synchronous written text and other facilities provides a platform from which to critically re-assess text-oriented discourse analytical methods in studies researching computer-supported language interactions (Herring, 2004). As early as the mid-90s, collaborative virtual learning environments were investigated through CA (Bowers, Pycock, O’Brian, 1996) but the aim was praxeological, concentrating on the identification of communicative shortcomings in order to advise on future technical refinements.

In the present paper, the aim is, instead, to identify methods for the analysis of language learner conversations in such environments so as to better understand how to promote multimodal conversation as a legitimate learning activity of the electronically literate. We do not endorse the view that technologies are a mere support for conversational activity, the script of which is then decoded through traditional language-centered methodologies such as discourse or conversation analysis. Instead we look upon technologies as mediating the social event that is the conversational process. In our work, ‘mediation’ is a construct with roots in the sociocultural theory of learning, and particularly in the work of Vygotsky (1978), Leontiev (1981) and Wertsch (1991). The sociocultural model of learning, as represented in the work of these authors, stresses the central role of interaction. Human learning is shaped within interactions involving mediational tools such as: other people, language, cultural assumptions, social institutions, technology, and the spatial and temporal characteristics within which the learning situation is played out. The shaping that takes place through these mediational tools

is iterative: the tools help create the learning, and in turn the learner shapes these tools, which further shape the learning and so on. For example early blog writers were diarists but the activity of bloggers in creating links has been so productive that sites like *Technorati* ([www.technorati.com](http://www.technorati.com)) have arisen, using the bloggers' links to index the interconnections between bloggers, who in turn use *Technorati* to create new links. By 'mediation', in this chapter, we refer to this mutual shaping. First, we establish the need for an understanding of what learners are doing when they converse via multimodal electronic systems. We then discuss the applicability of CA to multimodal conversations looking to identify and to bring together elements of a methodology that address these behaviours as a situated whole. Finally, drawing from communication theory and social semiotics, we identify three frameworks with the potential to answer questions raised by our data.

## 2. Learner conversations online and their analysis

### 2. 1. *The role played by conversation in language learning*

In language teaching, conversations are seen as beneficial to learners for at least three reasons. Firstly, they have been part of the communicative model of language teaching for half a century, often in the guise of conversation classes, in which learners, in the safety of the classroom, are invited to experience the 'pressure of conversation' (Cook, 1996 [1991]: 61) that it is assumed they will face when called upon to talk to native speakers in the 'target' country. Secondly, work carried out since the mid-80s within Second Language Acquisition theory has established that, providing they are structured so as to require participants to negotiate meaning, conversations promote socio-cognitive progress (Gass, Varonis, 1995; Long, 1998; Gass, Mackey, Pica, 1998; Chapelle, 2003). Thirdly, for language educationalists such as Belz, Thorne, (2005) and O'Dowd, (2006) who have more recently researched intercultural aspects of telecollaborative online projects (i.e. those projects which link groups at different institutions for example across the Atlantic), conversations are a means whereby learners from different cultures construct their knowledge of the conversational strategies of the target culture while developing their skills as interactants.

To attain these diverse learning outcomes, conversations have to satisfy certain criteria. They need to be part of a constraining task (for example, a group of learners might be asked to reach a negotiated consensus). However, to help learners cope with the spontaneity of speech, conversations also have to offer them sufficient time and freedom to develop new topics or to attend to unexpected conversational moves made by their discourse partners. Designing a model for conversation to be conducted via a single-channel technology such as the telephone is thus a challenging enough pedagogical project. Designing such a model is made more complex by the co-presence of different channels in systems such as voice-over-Internet groupware or virtual worlds (i.e. systems which allow oral and written communication in real time as well as the collaborative creation of images, and in the case of virtual worlds, the adoption of an identity via an avatar which is able to move or fly, allowing users to see the environment as if out of the avatar's eyes).

In our analysis (Lamy, 2006: 260) of a *WebCT* course, involving both written text (asynchronous and synchronous) and audio (asynchronous only), we observed that pre-

scribing the use of specific communication channels for particular discursive goals was ineffective, because users re-appropriated modalities in their own way, successfully developing conversations in the various phases of the task and in 'spaces' in the learning environment not originally designed for this purpose. For example, although we had structured the asynchronous forum into separate 'threads' for task-completion (in this case choosing fictitious identities) and for social interaction, one group quickly moved from a prescribed discussion of their fictitious surnames to a sustained conversation about their real surnames and other names in their families, making no attempt to change threads to do so. Building on this experience, we approached the setting up and delivery of the conversation-based audio-graphics course described below with two sets of expectations in mind: *pedagogically*, that users would appropriate channels and modalities for their own discursive purposes and *epistemologically*, that learners' unconstrained conversational behaviours across channels and modalities would provide us with opportunities to understand how they achieved their discourse aims. We refrained from any attempts to guide students in their use of the online environment. Our concern was with appraising the potential of CA when accounting for the real-time electronically-mediated conversational moves across different modalities.

## 2. 2. *Analysing conversations in the electronic medium*

Arising out of sociological work on everyday face-to-face interactions pioneered by Goffman (1967), CA was applied as early as the mid-70s to everyday conversations in a technologically-mediated setting, the telephone (Sacks, Schegloff, Jefferson, 1974). It was later extended to educational conversations in face-to-face situations (Wegerif, Mercer, Dawe, 1998). Wagner summarises the aims of CA as follows: "It is – as part of its ethnomethodological heritage – interested in describing how social order is produced in interaction" (Wagner, 1996: 232). In other words, CA identifies how interactants solve their conversational problems, seen as components of social activity. Our own object of study is both technology-mediated and learning-oriented, so for our participants, 'solving a conversational problem' may be expected to involve mediation issues about the integrated deployment of discursive and technical resources. Data collection should, therefore, be organized in such a way that the fullest account of users' interaction with their machines and their discourse partners can be created. In this respect, we have known since a study by Garcia and Jacobs (1999) that, providing the corpus contains video recordings of the users' individual computer screens as they process their input (i.e. as they type, delete, hesitate, retype etc.), CA may be seen as an appropriate analytical tool for analysing linguistic data collected from written text-chat. Garcia and Jacobs advocate that "[r]esearchers who wish to study CMC would do well to first understand the constraints – and the advantages – of the medium and to 'see' as completely as possible the process of the interaction" (Garcia, Jacobs, 1999: 363). This warning led to our combining computer-generated audio recordings of our subjects' oral exchanges with screen captures of the sessions. One advantage of the environment that we recorded in this way is that every keystroke is recorded on the shared word processor, and its author identified. Thus, as recommended by Garcia and Jacobs, as far as the word processor is concerned (though not in the text-chat) our screen videos reveal the input process.

### 3. CA and synchronous multimodality

Among the tenets of CA, two are pertinent to the data presented below. Like all CA principles, they are as productive – of conversational material – when participants abide by them as when they breach them. In a simplified form (see Wooffitt, 2005 for an in-depth presentation of CA) they can be summed up as follows: (a) turns of speech alternate and interlink. The basic principle is known as ‘conditional relevance’: given a question, expect an answer; given an apology, expect an acknowledgement; given a topic, expect that it will be pursued; (b) face-saving is always important in a conversation, because conversants are in a relationship of perpetually converging and diverging interests with their fellow-conversants. You may try to save your own face or to protect others’. It is also possible to threaten one’s own or others’ faces for strategic reasons.

In computer-assisted multimodal communication settings, discourse partners use many semiotic systems, including linguistic systems (written and spoken), iconic and symbolic systems. They may use these systems in rapid succession (type then draw), quasi-simultaneously (speak while hitting the ‘send’ button to post a written message) or choose among systems to make meaning in particular ways. For example, a user may close a conversation by typing ‘Bye for now’, by clicking a specific button, or by announcing their withdrawal orally. Different semiotic systems are also involved when responding to such a move. For example, if, as a user of a synchronous environment, I signal my impending departure orally, my discourse partners will receive aural input, made up of linguistic and non-linguistic information, e.g. a phone ringing in my home will suggest reasons for terminating my side of the conversation. Or I may type a valedictory message into a box and send it. In this case, my message will be displayed on the shared screens and will remain part of the ongoing conversation even after I have disappeared from the environment. Or again I might avail myself of the system’s symbol-based tools. For instance, some systems offer *telepresence indicators*, which effect symbolic changes to the objects on the screen, such as greying-out my name, or fad-

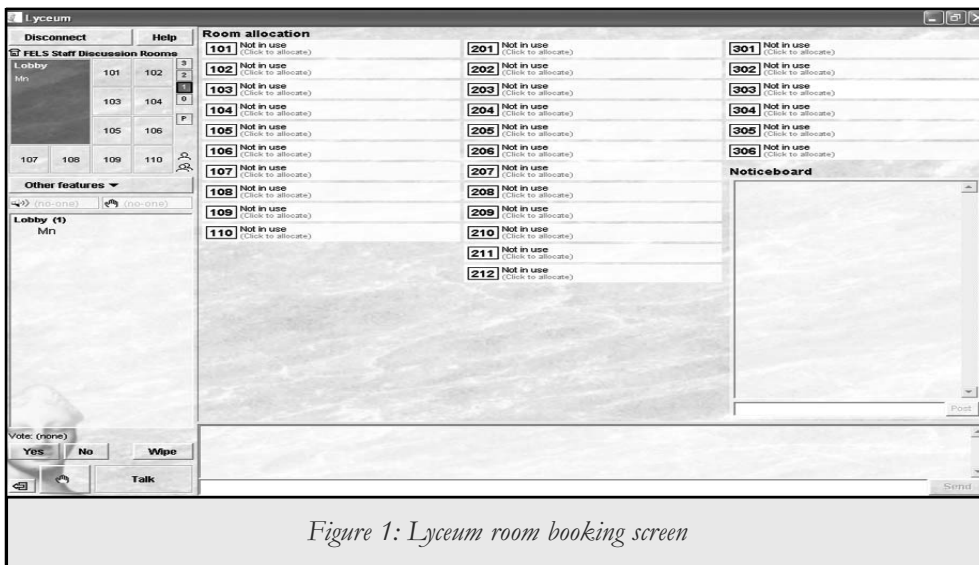


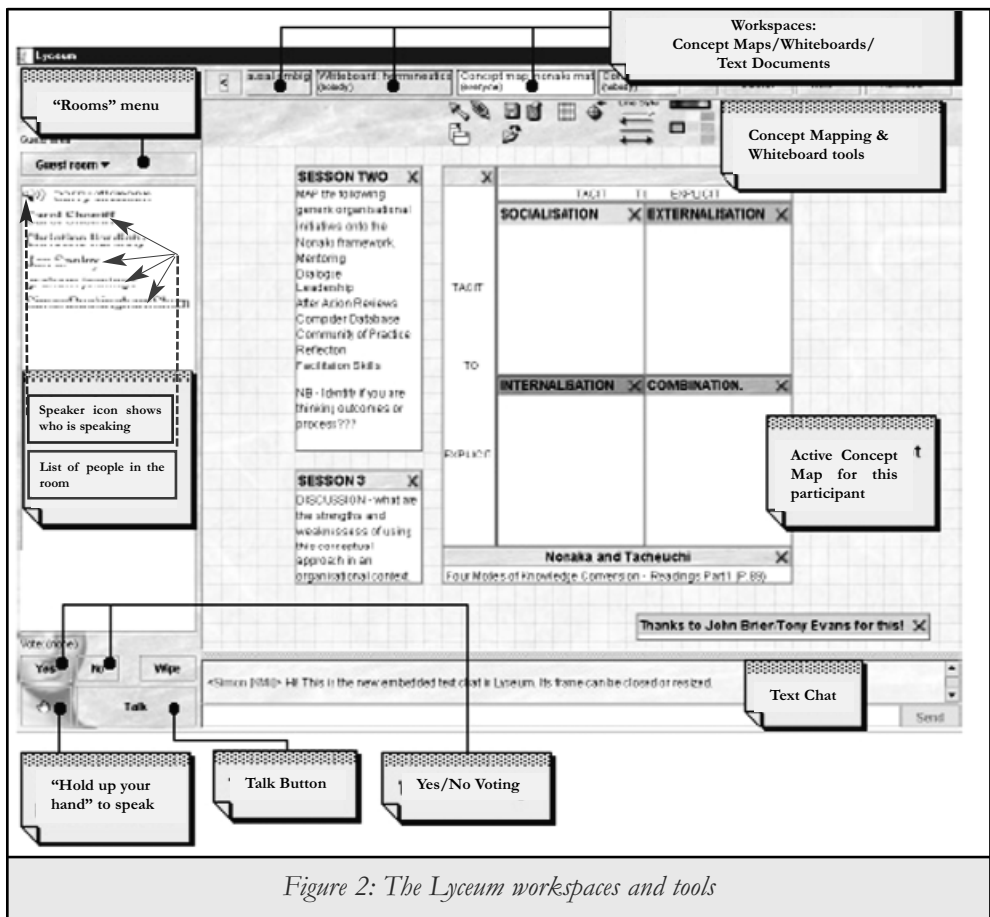
Figure 1: Lyceum room booking screen

ing-out my photo. Choosing to leave the conversation by activating one of these is a strategy likely to have particular effects on the ongoing conversation, distinct from the effect of disconnecting altogether (in which case my name or image would vanish), or from the effect of leaving all connections ostensibly active but walking away from my computer for the rest of the session. In environments with sophisticated meaning-making devices, such as webcam images of users, animations, translucent frames etc., the combinations and interlinking of semiotic systems and effects can become extremely complex. In the rest of this paper we seek theoretical support when attempting to grasp this complex reality, by studying three brief conversations carried out through the *Lyceum* audiographic groupware system as part of an English-for-Special-Purposes project. The next section briefly describes the system and the pedagogical setting, then introduces the three data sets.

#### 4. Analysis of three brief multi modal conversations

##### 4.1 The CoPéAs project: environment, population and pedagogical setting

*Lyceum* is a synchronous audiographic groupware system developed within the UK Open University, designed to facilitate distance tutorials. Figure 1 shows the home screen that



participants see just before they ‘join’ their group by clicking on the relevant room number. While technically the UK Open University’s servers may accommodate hundreds of connections at a time, the logistics of turn-taking is such that the optimal number of participants is 12 to 15 per tutorial room. However, rooms can be booked in a click, and the subdividing of plenary groups into separate sub-groups and back again is easy and almost instantaneous. Within each room, different tools and shared spaces can be displayed and collaboratively used. Figure 2 on the previous page shows the rooms-based facilities that *Lyceum* offers in support of flexible and dynamic grouping configurations. A turn-taking management system based on clickable icons is another feature of *Lyceum* that has relevance for our study.

As shown in Figure 3, a raised hand icon symbolises an intention to take a turn, while a loudhailer icon indicates who is speaking. However, protocols for managing turns of speech are left up to individual groups, as there is no privileged turn-allocator (all participants have the same permissions), nor is there any technical obstacle to ignoring the turn-taking tool altogether (although in practice no more than three people are recommended to speak at once, because of aural overload on listeners).

The extracts discussed below originate from a project concerned with teacher-learner communication in an audio-synchronous environment. This project, known as *CoPéAs* (*Communication Pédagogique et environnements orientés Audio-synchrones*), ran as a partnership between the UK Open University and the Université de Franche-Comté in France in the spring of 2005. Sixteen French-speaking students studying for a Professional Masters in Open and Distance Teaching (ODT) worked in two groups of eight, connecting from their homes in various parts of France. Each group had a UK-based English native speaker tutor connecting from the UK. The groups met during 10 sessions of over an hour each. The course had a dual objective, linguistic and vocational, which was the development of competences in ODT through spoken and written English. The less proficient group (false beginners with wide internal variations in knowledge) provided the extracts discussed below.

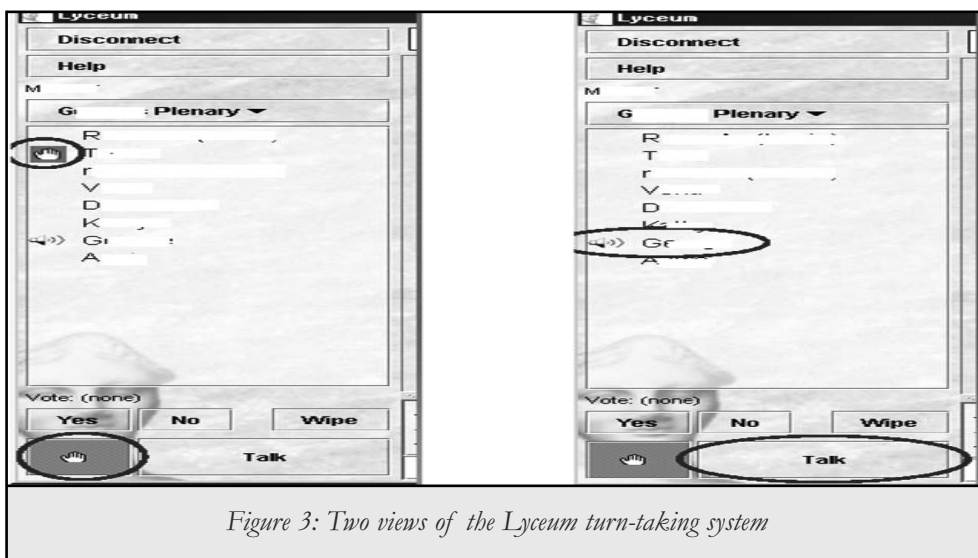


Figure 3: Two views of the Lyceum turn-taking system

The *Lyceum* environment and the *CoPéAS* project are presented in detail in Vetter, Chanier (2006). *Lyceum* supports semiotic resources for constructing discourse in interaction, such as natural language in its written and spoken forms, as well as visual resources such as icons, images, colours and shapes. Modalities in this chapter are understood as the material through which the conditions for human interaction are created when humans use *Lyceum*. Just as, in pre-mechanical written interaction, physiological material (bodies writing) and physical equipment (pen and paper) stand in a relationship of mediation with the semiotic resource ‘written language’, with which it co-constitutes the interaction, so in the computer-based situation, does technological material. In *Lyceum* such material includes keys to be pressed for text creation and entry, pads or mouse attachments to be tapped or clicked for effecting various actions, headsets to be mounted and enabled for oral communication. Thus henceforth ‘a modality’ refers to one such technological feature, in its relationship with a particular semiotic resource for example a visual signifier such as a frame, button, symbol

User	T	Time	Audio Transcript	Chat log	Tick	Doc
H	1	28:17	vocal er bulaire er technical vocabulaire je sais pas comment on dit			
C	2	28:25	vocubula vocabulary <laughter>			
C	3	28:28				C writes: vocabulary learning
A	4	28:33	er ah yes maybe er			
C	5	28:36	do you think it's er ok, H?			
H	6	28:46		technical vocabulary		
H	7	28:50	er technical er vocabulary er technical vocabulary			
C	8	28:51				C deletes previous and inserts: to learn technical vocabulary
H	9	29:02			✓	

When the first extract begins, the learners are engaged in an evaluation-type conversation. Three students, H, C and A are communicating orally negotiating an agreed heading, summing up what they think they mostly learned during the course. The heading then needs to be typed on the shared document screen by one of the group. In *Extract 1* (Table 1), H twice tries to pronounce the phrase *technical vocabulary* (at T1 and T7) but stutters and hesitates each time. As C attempts to settle the answer by typing *vocabulary learning* in the shared document (T3) and eliciting H's approval of this formulation (T5), H holds to his original wording by writing it in the chat window (T6). This results in C modifying (T8) what he had typed on the shared document, which H finally approves (T9) by clicking the Tick icon. Arrows show topic-maintenance by H across modalities. To compensate for his articulatory problems in English, H deploys an alternative conversational strategy that takes advantage of the environment's multimodality. His strategy cannot be made visible through conversation analysis conducted in its classical form since CA relies on the sequentiality principle, and the conversation, if read vertically down the second column of Table 1, has no sequencing that could be sensibly interpreted. Sequentiality is in fact present, but can only be detected by analysing the four rightmost columns.

Table 1: *Extract 1: reinterpreting sequentiality*

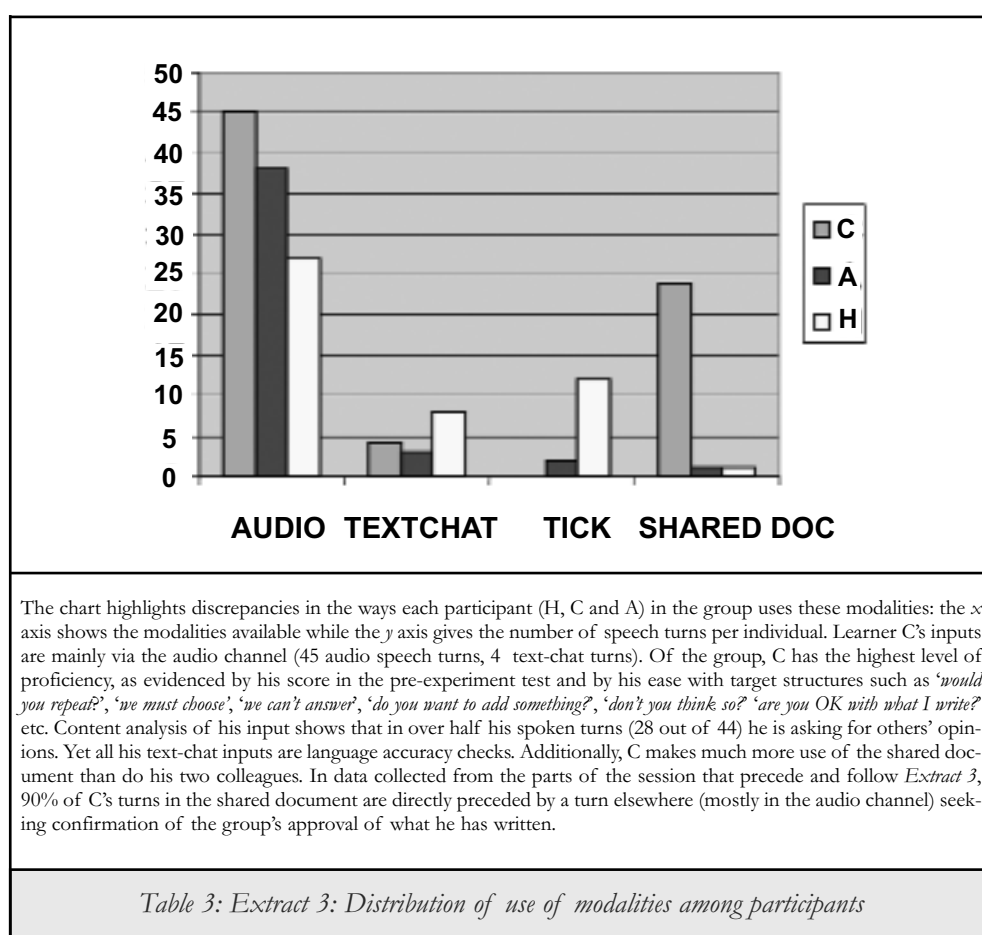
User	T	Time	Audio Transcript	Chat log
A	1	30:36	Er I do my presentation er my name is er A donc I come from er I live er at Sète er a town near Montpellier and er I come from er west [inaudible] but I er I was born in Toulon er in a town er south er of France	
P	2	30:57		A, I don't read* you [*P uses this verb to mean 'hear' perhaps influenced by war film cliché]
Tutor	3	31:13	ok thanks er A er so C is C ... if C is there would you like to talk a bit about yourself C?	
[...] C answers Tutor's question (in audio) at length [...]				
A	4	36:14		P, I don't write
P	5	36:59		no, A, you need to put the volume of your microphone more high
C	6	37:08	er yes <laughter>	
P	7	37:12		I think...
Tutor	8	37:14	ok and do they install they install motors as well as in in in er in houses or that kind of thing?	
G	9	37:45		I don't understand
C	10	37:49	can you repeat your question please?	
P	11	37:50		me too
A	12	39:41		sorry, P, but I can't
Tutor	13	39:54	sorry did I did you say fourteen to seventeen years old?	
P	14	40:03		ok ok
<p>Just before the start of Extract 2, the tutor had launched the conversation by asking the members of the group (A, P, G and C) to introduce themselves orally. The extract in Table 2 on the facing page shows the conversation proceeding orally, with some input typed into the text chat window (highlighted in grey-blue), while two of the learners (A and P) carry out a completely different conversation in text chat (highlighted in green), concerned with an auditory difficulty. In the recording which was used for the transcription, A speaks quite audibly but, possibly due to her home set-up or for server-related reasons, P could not hear him. So A and P run a conversation in parallel with the main conversation initiated by the tutor. They use a second modality (chat) to construct a dialogue on a different theme, without apparently disturbing the 'main' conversation which is proceeding in the audio modality. Thus, their attitude may be seen as non-transgressive in terms of saving the tutor's face. In a non-virtual classroom, such a move would probably have been seen as face-threatening for the tutor, who would have taken steps to stop it, had it persisted for more than a few seconds. The face-saving principle helps explain A and P's conversational moves. However, the example shows how this principle needs to be re-interpreted for this online multimodal communication scenario.</p>				
<i>Table 2: Extract 2: reinterpreting the face-saving principle</i>				



or icon, as potential or realised co-constituent of a human interaction. A particular semi-otic resource may therefore be associated with several modalities, or with a single one (as for spoken language and the audio modality). In the following data, we are concerned with two semiotic systems (written language and spoken language) and four modalities (audio, chat, shared document and voting or 'Tick' system).

#### 4.2. The data

The three extracts occur in the last seven minutes of the last session in the course. Both Table 1 and Table 2 give simplified versions of a tabular transcription of the screen videos. Reading the top row of Table 1 from left to right, the transcripts show: participants' initials (*User*), the speech turns (*T*, characterising all inputs, and not solely those in 'speech'), chronological sequences (*Time*) and data collected from the four modalities (in the four rightmost columns). In the last example, Table 3, we see how learner C uses the four modalities at his disposal: audio, text-chat, Tick icon and shared document. We hypothesize that C uses the text-chat modality when he is checking the accuracy of the lexicogrammatical forms



he is using rather than moving through the conversational agenda of the class. He prefers to progress through the class agenda via the shared document modality, but only after obtaining consensus from his peers orally. This pattern of use could relate to two different yet converging factors: C may be specialising particular discourse aims to particular screen elements, and his representation of himself as a communicator may also play a role in which of these he uses. For example, in relation to the elements in the screen layout, analysts of his conversational strategies would need to assess the degree of salience of the text-chat window tucked away unobtrusively (and indicated in the bottom right-hand corner of Figure 2 by the *Text Chat* label) in contrast to the intrusive quality of the sound coming through each group member's earphones. As for self-representation, we may posit that face-saving issues are involved: for example C might be prepared to give an image of himself as a confident language speaker in the audio channel while specialising the chat window for more risk-taking face-threatening activities, such as asking for help with English forms. A similar explanation might be offered for his use of the shared document: the visual organisation of the screen when that document is uploaded to it (see Figure 4), the central disposition of that document and its status as the 'official' record of the group's collaboratively negotiated view may explain both his self-appointed guardianship of it and his diffidence in committing material to it, unsupported by his peers.

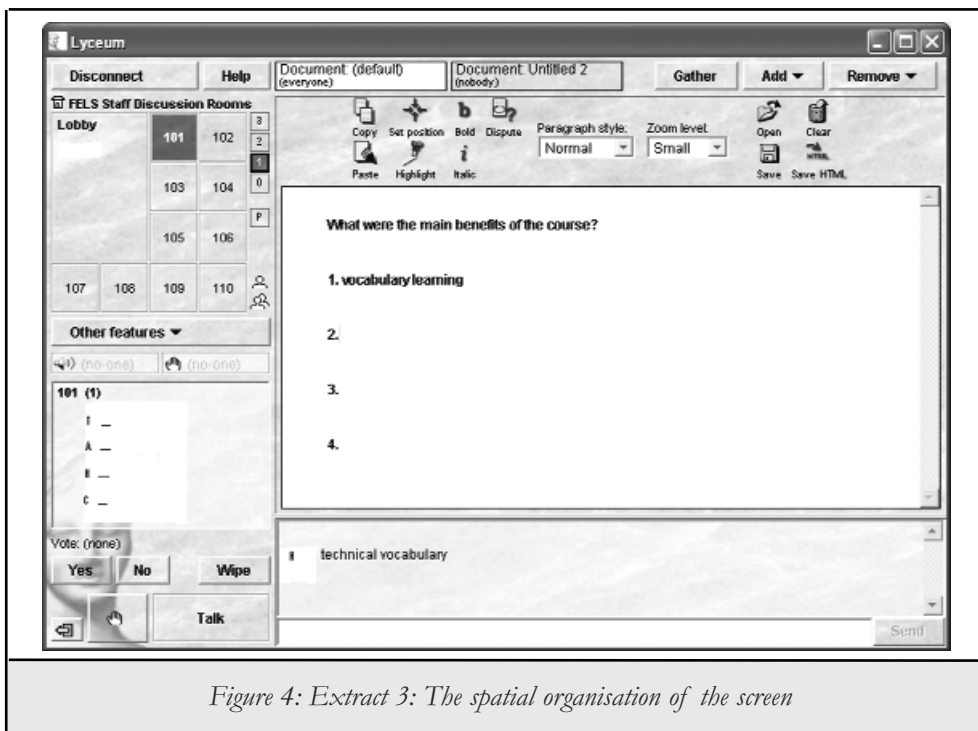


Figure 4: Extract 3: The spatial organisation of the screen

## 5. Discussion

In this section and on the basis of the findings from the extracts presented earlier we suggest that while CA remains a useful approach to the understanding of meaning-making in real-time online multimodal settings, it needs to be rearticulated. Our suggestion is that such reframing can be done in the light of three theoretical frameworks drawing from affordance theory as developed by Gibson (1979), social semiotics and geosemiotics respectively. Our reading of the data showed that two of the principles of traditional CA, sequentiality (*Extract 1*) and face management (*Extract 2*) could be used to characterise conversational moves within electronically-mediated multimodal conversations, albeit redefined, because when more than one modality was available, both sequentiality and face management operated transmodally. Hutchby's work is, we suggest, relevant in this respect.

### 5. 1. Hutchby and communicative affordances

Technology, Hutchby (2001) reminds us through the example of the early history of telephone communication, exists in a reciprocal relationship with its users. The telephone was originally sold as an instrument for transacting business or for exchanging practical information. Early marketing stressed this functional use to the extent of advising subscribers to wait until late at night if they really needed to use the machine for the purpose of personal chat. However users were not persuaded and appropriated the telephone as a medium for family or other intimate conversations, forcing the telephone companies to review their marketing strategies:

[...], while designers may be said to have some control over the features they design into an artefact, and while they may have some idea about the range of uses to which the artefact should be put, they have little control over the artefact's communicative affordances – over the range of things it turns out to enable people to do.

(Hutchby, 2001: 123)

We witnessed this mechanism at work in the *CoPéAs* extracts, where learners appropriated modalities in diverse ways. For example, the text-chat resource was re-appropriated in the contexts of different conversations in order to support control over content (*Extract 1*), to provide technical assistance (*Extract 2*) and to enact face management (*Extract 3*). We also observed that individual participants engaged with different communicative affordances to satisfy identical communicative needs. For example legitimisation of self as a turn-taker (and therefore control over the conversation) was achieved via moves from audio to text-chat then to the Tick icon by H (in *Extract 1*), but for C (in *Extract 3*) by toggling between the audio channel and the shared document. The construct of communicative affordance – as described by Hutchby in the last line of the quotation above – also helps to understand the different ways in which human actors in different technological settings solve a single problem such as (to stay with the example of the telephone) answering a landline call, a mobile phone or a computer bleep from an Internet-telephony system. The same conversational problem is involved in each case (how to respond to a conversational invitation from a remote caller), but different discursive solutions are appropriate in each case. For instance, a person called on a landline without user ID

will in most ordinary circumstances pick up the handset and initiate their side of the conversation by uttering a conventional query (e.g. *bello, allô, pronto* etc.) in a rising tone followed by a pause, in the expectation that the caller will identify themselves. With user ID, or in Internet telephony where the caller's name and sometimes photograph appears on screen, the person called is likely to respond instead with a greeting usually followed by the caller's name. A pause, if provided by the person called, will not be heard by the caller as an invitation to identify him/herself; instead it will require interpretation and negotiation so that the conversation may progress. Although a close textual analysis of our learners' input is not the focus of the present methodological study, we follow Hutchby in maintaining that:

The detailed ways in which people have taken up the affordance for verbal intimacy across distance and shaped it for communicative ends can only be revealed once we move beyond looking at the cultural meanings or representations of the telephone as a 'technological artefact' and observe telephone talk itself.

(Hutchby, 2001: 31)

In this view, CA (which claims to be able to 'observe talk itself') continues to be central to understanding conversations in electronically-mediated multimodal settings. However, only by paying attention to communicative affordances will we be able to carry out the necessary work of re-interpreting its principles in terms of their transmodal manifestations and discursive adaptations.

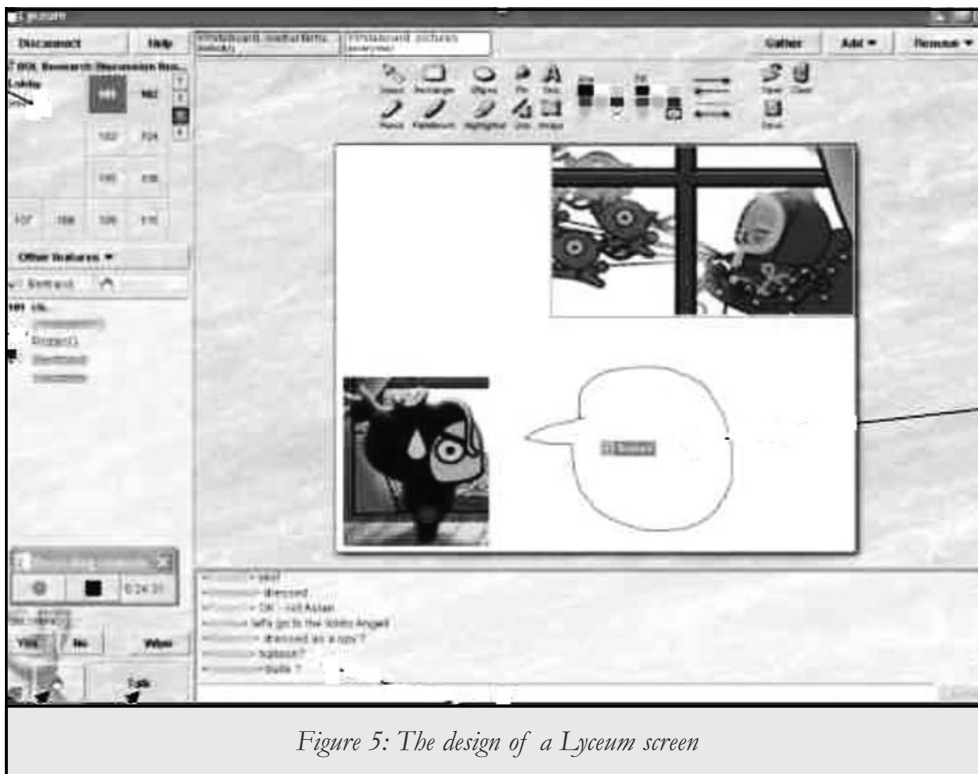


Figure 5: The design of a Lyceum screen

### 5. 2. Kress and van Leeuwen: design and discourse

The materiality of the environment impacts on the dynamics of conversations. In *Extract 2*, two conversations were progressing in parallel. It is not accidental that the tutor-initiated one was carried out orally while the students' exchange about sound levels was conducted in text-chat: all we need to do to become persuaded of this is to mentally invert the modal choices and imagine that the tutor led his tutorial via postings in the text-chat while students talked about other topics in the audio channel. It is unlikely that the group would accept such a position for the tutor, and we draw from social semiotics to help explain why.

Kress, van Leeuwen's (2001) work on the semiotics of multimodal pages in newspapers and books identifies the four dimensions of structuring for such artefacts: discourse, production, dissemination and design. While we believe that all four are relevant to understanding meaning-making in electronic learning environments (see for example a discussion of the impact of dissemination on learner progress in Lamy, 2004: 523-524), for the purpose of the current discussion we concentrate on one of them, the dimension of design. The designers of the software that our learners used can be appositely compared with the architect in Kress and van Leeuwen's explanation of the relationship between design and discourse:

An architect, for instance, designs (but does not build) a house or a block of apartments. The discourse provides a certain view of how houses are lived in the way they do, and arguments which critique or defend this way of life. The design of the house then conceptualises how to give shape to this discourse in the form of a house, or a type of apartment.

(Kress, van Leeuwen, 2001:6)

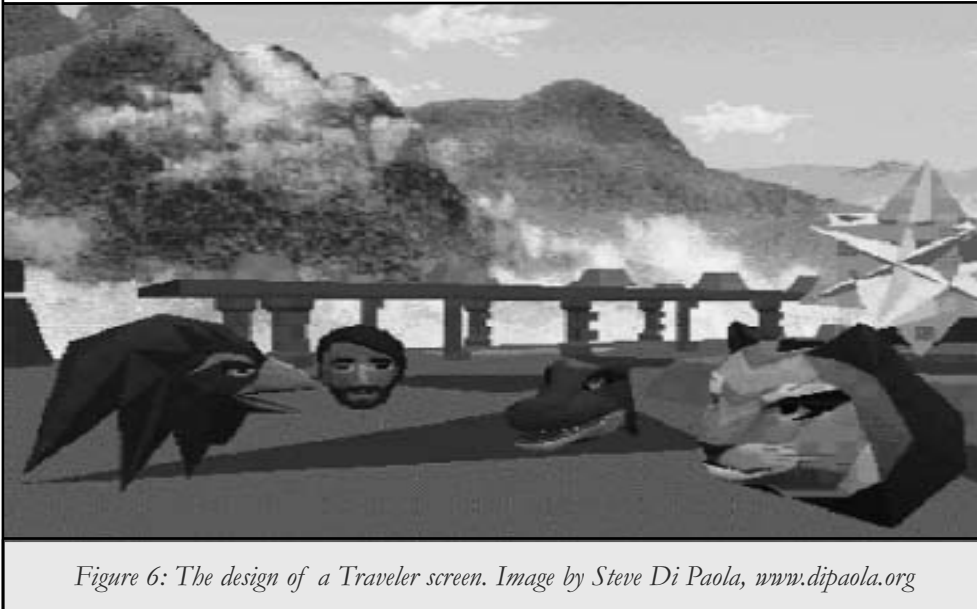


Figure 6: The design of a Traveler screen. Image by Steve Di Paola, [www.dipaola.org](http://www.dipaola.org)

The relevance of this architectural example will become clear as we look at two very different designs of environments used for language learning, *Lyceum* and *Traveler*, through comments made by learners. *Lyceum* (Figure 5), as we have seen, is an academic tool designed to look like a university campus building. *Traveler* (Figure 6) is an avatar-based system with a ‘fantasy’ feel inherited from the world of games. Just as the architect has provided a shape for the cultural discourses of human habitats, so the designers of electronic environments and virtual worlds conceptualise the form that the environments take on the screen into interpretable signs. For example here are responses from two of our users (for *Lyceum*) and two of Örnberg Berglund’s (2005) users (for *Traveler*), upon being asked about their feelings of ‘presence’ when online:

*Lyceum* user 1: ‘Quand le prof rentre dans la salle, cela ne dérange pas. Je sais pas comment l’expliquer’. (When the teacher enters the room, it’s not intrusive. I don’t know how to explain it).

*Lyceum* user 2: ‘Le style du prof joue, mais le fait qu’il est invisible, il ne peut pas s’imposer de la même façon qu’en présentiel’. (The teacher’s style is a factor but the fact that he’s invisible, he can’t impose himself in the same way as in face-to-face).

*Traveler* user 1: ‘It took me to another world and was a real adrenaline buzz. It was on my screen and I was conscious of it always, but I was definitely virtually gone from my usual habitat’.

*Traveler* user 2: ‘I’m always immersed.[...] It doesn’t matter that the environment is artificial. [...] I think of the place as real’.

Whereas these two *Traveler* users produce a discourse of emotions and escapism, the discourse of *Lyceum* users reflects school-like representations of a particular type: teacher-led classes. We make two comments here. Firstly, although these differences in perception may not be surprising given the strongly contrasted visual identities of Figures 5 and 6, the question is whether two groups using these environments for achieving the same language-learning objectives might have very different types of conversations in each environment. The second observation relates to designers and the uses made of their designs. *Lyceum*’s design was underpinned by a democratic and participative pedagogical posture: “We have imposed minimal technical constraints on ‘floor control’: anyone can speak anytime” (Buckingham-Shum, Marshall, Brier, Evans, 2001: 4). Yet the users’ comments show a pre-occupation with teacher control. It is likely that this is part of their pre-existing non-virtual educational culture, in which case the question can be asked: to what extent and in what ways can the design features of interactive learning environments transform the users’ representations of self? The answer to this question is another determiner of sense-making in these environments. Finally, regarding the device which introduced this section, i.e. the proposal that the *Lyceum* tutor could conduct core tutorial business in the chat box while the students conversed orally, evidence from *Lyceum* users’ perceptions supports the view that the system’s design provides a shape for the cultural discourses of traditional teacher-centered classrooms. But based on social semiotics’ understanding of design, there is no in-principle reason why other types of design could not work to support other cultural discourses, producing distinct types of conversations.

### 5. 3. Scollon and Scollon and perceptions of space

While Scollon, Scollon (2003) acknowledge the importance of Kress and van Leeuwen's design dimension, they re-interpret and re-inforce it in order to account for interaction, in a way which we have found useful for understanding how our users perform conversations by choosing among the different spaces offered to them within the *Lyceum* interface. Scollon and Scollon call their framework 'geosemiotics' and define it thus:

Geosemiotics is the study of meaning systems by which language is located in the material world. This includes not just the location of words on the page you are reading now but also the location of the book in your hands and your location as you stand or sit reading this

(Scollon, Scollon, 2003: x-xi)

The authors structure geosemiotics into three sub-sets: the interaction order, visual semiotics (on which we will not elaborate here, as this concept comes close to Kress and van Leeuwen's notion of design mentioned earlier) and space semiotics. The understanding of space that is of interest to us in our study of learners using a virtual computer environment is predicated on each of these three sub-sets.

The interaction order provides a construct for understanding how individuals perceive the interactional value of the space they choose to use. In their description of the interaction order, the authors include perceptual spaces and interpersonal distances. Dominant perceptual spaces are visual and auditory ('less noticed' ones are olfactory, thermal and tactile). The addition of the construct of interpersonal distances – as a scale of values inspired by Hall's (1969) work on proxemics – allows geosemiotics to ask questions about the relationship between space, sound and socialisation. For example, the auditory space which I perceive and my perceived intimacy or distance with the individual vocalising the sound that I am hearing, together form the semiotic resource by which I embody meanings. Applying this framework to *CoPéAs* participants, in particular to the parallel conversation mechanism in *Extract 2* and to the multimodal preferences of the learner in *Extract 3*, the question becomes: how do they co-construct interpersonal values (intimate, personal, social, public) into conversations which proceed simultaneously through visual spaces of varying salience and through an auditory space defined by the spatially and tactilely intimate device of an earpiece or headset?

Space semiotics, in Scollon and Scollon's words, is the most fundamental part of geosemiotics, because it asks "Where in the world is the sign or image located?" and because it aims to account for "any aspect of the meaning that is predicated on the *placement* of the sign in the material world" (2003: 146, our italics). The distinction between well-rehearsed debates within visual semiotics on the subversive placement of images for artistic purposes (e.g. Warhol soup cans), and the focus of space semiotics, is that the latter is looking at the material world as a whole, and not simply at materials used for the bearing of signs, such as paper, canvas or brick. In terms of multimodal electronic environments, space semiotics provides the basis for asking questions such as: how do users decode and encode meanings in a material situation involving their computer and its various peripherals (keyboard, mouse or keypad, webcam) as well as other stimuli around them (possibly another computer, a video screen, a person physically present, who is talking, writing or drawing, etc.)?

To put this in semiotic (rather than in multitasking) terms, consider the difference in meaning-making between two situations that the author recently experienced. One was when she created a blog at home using her desktop machine and sheets of paper to help sketch out the blog's design, the other when she created a blog using a laptop with a *wifi* connection while attending a talk given by a speaker on the subject of creating blogs. To further interrogate the second situation, what are the semiotic implications of the two possible scenarios that follow: in scenario one the laptop user was simply carrying out the lecturer's instructions; in scenario two the lecturer was unaware of what the laptop user was doing, until she publicly revealed it, as part of her contribution to an end-of-lecture debate on the positive and negative socio-pedagogical consequences of *wifi* technology.

## 6. Conclusion

As the methodological approach put forward in this paper develops, we are aware of stepping into a territory that is just beginning to be charted for some environments (for example as far as multiplayer immersion games are concerned, see Lemke, 2006) but not yet for others, such as educational virtual environments. Although Jewitt, Kress (2003) have used social semiotics in their work in a field sometimes called 'multimedia semiotics' to lay the foundation for an understanding of learning via multimodal texts, and more recently via educational software (Jewitt, 2004), the interaction dimension is not part of such work to date. However emerging research on Finnish sign language speakers in multimodal environments, such as McCambridge (2006), has independently arrived at the conclusion that similar theoretical frameworks (to those that we discussed in this paper) can be of use. This convergence will be further appraised as the Finnish work becomes available.

Thus, we see a useful research agenda emerging: to test out, with a large volume of learner interaction data collected from multimodal environments, the methodological claims made in this paper according to which such data can be best analysed through the synergistic use of conversation analysis, social semiotics and geosemiotics. However, four of the authors that have provided the frameworks discussed in this paper agree on a need to approach this task not with a grand vision, but in a practical manner, drawing on the theoretical models at the point of need. This is how Scollon and Scollon express this consensus:

We don't believe we can assume that there are general, grand, and overarching semiotic systems. Certainly the field is too new to try to establish such systems. We prefer to follow Kress and van Leeuwen's preference for thinking in terms of small systems of meaning interacting with each other.

(Scollon, Scollon, 2003: 157)

Based on his work with virtual games worlds, Lemke echoes this practical approach as he sketches out a research methodology which, we suggest, is applicable to educational virtual worlds. Accordingly, we leave the last word to him:



Perhaps most importantly for research purposes, it is possible to create real-time, synchronized video and computer log records of monitor display, keystroke and mouse or joystick input, and user speech and action. It is possible in this way to follow user activity in entering the gameworld, acting and moving within the primary gameworld and among various subsidiary “screens” or auxiliary attentional spaces, communicating within and parallel to the gameworld action [...], and on leaving the gameworld. Ideally we would also like to observe how people integrate or cumulate in-game meaning-making activity and meanings made with out-of-game life activities and identities. [...] We would like to understand class, gender/sexuality, cultural and subcultural differences in which games people play, how, and why; the kinds of meanings they make and feelings they experience; and what persistent learning effects result. But we need to take such an ambitious agenda one step at a time.

(Lemke, 2006: 11.)

### **Primary textual sources**

<http://www.dipaola.org>

<http://www.informatik.uni-trier.de/~ley/db/conf/chi/chiC96.html>

<http://www.technorati.com/>

## Scientific works

### Online Conference Proceedings.

Bowers, John, Pycock, James, O'Brian, Jon (1996). "Talk and Embodiment in Collaborative Virtual Learning Environments". *Proceedings of the 1st Computer Human Interaction Conference, Computing Systems: Common Ground*. ACM Digital Library

### Printed works:

Belz, Julie, Thorne Stephen (eds.) (2005). *Internet-Mediated Intercultural Foreign Language Education*. Boston: Thomson Heinle.

Buckingham-Shum, Simon, Marshall, Samuel, Evans, Tony (2001). "Lyceum: Internet Voice Groupware for Distance Learning". In *Proceedings of the 1st First European Conference on Computer-Supported Collaborative Learning*. Maastricht: The Netherlands.

Chapelle, Carol (2004). "Technology and Second Language Learning: Expanding methods and agendas". *System*, 32: 593-601.

Cook, Vivian (1996 [1991]). *Second Language Teaching and Language Learning*. London: Arnold.

Garcia, Angela, Jacobs, Jennifer (1999). "The Eyes of the Beholder: Understanding the turn-taking system in quasi-synchronous computer-mediated communication". *Research on Language and Social Interaction*, 32, 4: 337-367.

Gass, Susan, Mackey, Alison, Pica, Teresa (1998). "The Role of Input and Interaction in Second Language Acquisition". *Modern Language Journal*, 82: 299-307.

Gass, Susan, Varonis, Evangeline (1994). "Input, Interaction, and Second Language Production". *Studies in Second Language Acquisition*, 16, 3: 283-302.

Gibson, James Jerome (1979). *The Ecological Approach to Visual Perceptions*. Boston: Houghton Mifflin.

Goffman, Erving (1967). *Interaction rituals: Essays on face-to-face behaviour*. New York: Random House.

Hall, Edward (1969). *The Hidden Dimension*. New York: Doubleday.

Herring, Susan (2004). "Computer-Mediated Discourse Analysis: An approach to researching online behavior". In Sasha Barab, Rob Kling and James Gray. (eds.), *Designing for Virtual Communities in the Service of Learning*. New York: Cambridge University Press, pp. 338-376.

Hutchby, Ian (2001). *Conversation and Technology: From the telephone to the Internet*. Cambridge: Polity.

Jewitt, Carey, Kress, Gunther (2003). *Multimodal Literacy*. Bern: Peter Lang Publishing.

Jewitt, Carey (2004). "Multimodality and New Communication Technologies". In Philip LeVine and Ron Scollon. (eds.), *Discourse and Technology: Multimodal discourse analysis*. Washington DC: Georgetown University Press, pp. 184-195.

Jewitt, Carey (2005). *Technology, Literacy, Learning: A multimodal approach*. London: Taylor and Francis.

Kress, Gunther, Van Leeuwen, Theo (2001). *Multimodal Discourse: the modes and media of contemporary communication*. London: Arnold.

Lamy, Marie-Noëlle (2006) "Interactive Task Design: Metachat and the whole learner". In María del Pilar García Mayo. (ed.) *Investigating Tasks in Formal Language Learning*. Clevedon: Multilingual Matters.

Lemke, Jay (2006). "Towards critical Multimedia Literacy: Technology, research, and politics". In Michael McKenna, Linda Labbo, David Reinking and Ronald Kieffer (eds.). *International Handbook of Literacy and Technology* Volume III. Mahwah, New Jersey: Laurence Erlbaum Associates, pp. 3-14.

Leontiev, Alexis (1981). *Psychology and the Language-learning Process*. Oxford: Pergamon Press.

Long, Michael (1983). "Linguistic and Conversational Adjustments to Non-Native Speakers". *Studies in Second Language Acquisition*. 5, 2: 177-193.

McCambridge, Elina (2006). "Language Practices of Finnish Sign Language Speakers in Multimodal Environments". Paper presented at Eurocall 2006 Conference, Granada, Spain, 4 -7 September, 2006.

O'Dowd, Robert (2006). *Telecollaboration and the Development of Intercultural Communicative Competence*. Berlin, Langenscheidt.

Örnberg Berglund, Therese (2005) "Multimodality in a Three-dimensional Voice Chat. In *Proceedings of the 3rd Conference Multimodal Communication, Papers in Theoretical Linguistics*. Göteborg: Göteborg University.

- Sacks, Harvey, Schegloff, Emanuel, Jefferson, Gail (1974). "A simplest systematics for the organisation of turn-taking for conversation". *Language* 50: 696-735.
- Scollon, Ron, Scollon, Suzie Wong (2003). *Discourses in Place: Language in the material world*. London and New York: Routledge.
- Vetter, Anna and Chanier, Thierry (2006). "Supporting Oral Production for Professional Purposes in Synchronous Communication with Heterogenous Learners". *ReCALL* 18, 1: 5-23.
- Vygotsky, Lev Semionovitch (1978) *Mind in Society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wertsch, James V. (1991). *Voices of the Mind: A sociocultural approach to mediated action*. London: Harvester Wheatsheaf.
- Wegerif, Rupert, Mercer, Neil, Dawe, Louise (1998). "Software design to support discussion in the primary curriculum". *Journal of Computer Assisted Learning* 14: 199-211.
- Wooffitt, Robin (2005) *Conversation Analysis and Discourse Analysis: A comparative and critical introduction*. London: Sage.