ReCALL **20**(2): 162-182. 2008 © European Association for Computer Assisted Language Learning
DOI: 10.1017/S0958344008000426 Printed in the United Kingdom

Developing online multimodal verbal communication to enhance the writing process in an audio-graphic conferencing environment

MAUD CIEKANSKI AND THIERRY CHANIER
LASELDI, Université de Franche-Comté, 30, rue Mégevand,
25030 Besançon cedex, France
(*email: {maud.ciekanski, thierry.chanier}@univ-fcomte.fr*)

Abstract

Over the last decade, most studies in Computer-Mediated Communication (CMC) have highlighted how online synchronous learning environments implement a new literacy related to multimodal communication. The environment used in our experiment is based on a synchronous audio-graphic conferencing tool. This study concerns false beginners in an English for Specific Purposes (ESP) course, presenting a high degree of heterogeneity in their proficiency levels. A coding scheme was developed to translate the video data into user actions and speech acts that occurred in the various modalities of the system (aural, textchat, text editing, websites). The paper intends to shed further light on and increase our understanding of multimodal communication structures through learner participation and learning practices. On the basis of evidence from an ongoing research investigation into online CALL literacy, we identify how learners use different modalities to produce collectively a writing task, and how the multimodal learning interaction affects the learners' focus and engagement within the learning process. The adopted methodology combines a quantitative analysis of the learners' participation in a writing task with regard to the use of multimodal tools, and a qualitative analysis focusing on how the multimodal dimension of communication enhances language and learning strategies. By looking at the relationship between how the learning tasks are designed by tutors and how they are implemented by learners, that is to say taking into account the whole perception of multimodal communication for language learning purposes, we provide a framework for evaluating the potential of such an environment for language learning.

Keywords: multimodality, CMC, collaborative learning, writing process, audio-synchronous environment, learning strategies

1 Introduction

In his last review concerning the evolution of technology choice in the area of computer assisted language learning (CALL), Stockwell (2007: 113) shows how technology has moved on from CALL to computer-mediated communication (CMC) and computer-

supported collaborative learning (CSCL). This evolution concerns every language skill and area. Writing therefore is no longer perceived only in its personal dimension, but as an interactive process which may be mediated successfully by computers and groups of learners. Previous experiments, such as that of Dejean and Mangenot (2000), have shown successful collaboration between learners in front of the same computer in writing one text, and how the screen provided a convergent effect, facilitating the collaboration. These experiments have highlighted the importance of discussion about writing in order to help learners develop their writing learning awareness.

The recent development of synchronous online environments, integrating a large range of modes, has provided the opportunity to set up pedagogic and communication scenarios designed to enhance collaboration supported by a combination of modes and modalities. Three types of multimodal synchronous online environments can be listed. These environments refer to different kinds of multimodality and communication:

- an audio-synchronous environment integrating audio and textchat (verbal communication) (eg. Jepson, 2005);
- a videoconferencing environment integrating audio, video and textchat (verbal and non-verbal communication) (Andrews, 1993; McAndrew, 1996; Stevens & Altun, 2002; Wang, 2004);
- an audio-graphic conferencing environment integrating audio, graphics and textchat (verbal and non-verbal communication).

This paper reports on an Audio-Graphic Synchronous Environment (AGSE), which includes communication tools and shared editing tools. Previous recent studies analysed multimodal communication in AGSE. These studies take into account a wide range of dimensions: modes affordances (Hampel & Baber, 2003), task design (Hampel, 2006), oral communication (Lamy, 2006; Chanier *et al*., 2006), tutoring practices (Vetter, 2004; Hampel & Hauck, 2005), a multimodal communication model (Chanier & Vetter, 2006) and multimodal communication strategies (Jeannot *et al*., 2006). If communication, whether occurring face to face (F2F) or online, is always multimodal, in the online situation the computer medium has an impact not only on the types of modes involved, but also on the nature of what online communication is.

Our study follows on from the first observations made by Vetter and Chanier (2006) concerning the correspondence between modes defined as "semiotic resources which allow the simultaneous realisation of discourses and types of (inter)action" (Kress & van Leeuwen, 2001:21). Whereas the majority of previous investigations have been based on oral communication, we focus here on writing skills. This paper establishes an original methodological approach which aims at contributing to a better understanding of the characteristics of online multimodal verbal communication in AGSE. Additionally important is the development of the writing process, which is perceived as a social event and as a complex and procedural activity requiring not much instruction but action. This perspective will be of interest to language teachers who want to design online collaborative writing activities intertwined with synchronous communication, as well as to researchers who want to further investigate the relationships between writing and multimodal communication from the general view of online communication stressed by Lamy and Hampel (2007).

In section 2, we set the scene by presenting the learning situation, an ESP course for graduate students in a master's degree programme in open distance teaching (ODT), and the characteristics of the AGSE environment. Then, in section 3, we define the term 'multimodality' as related to our framework and emphasize three main modalities which support the writing process, i.e. textchat, audio, and actions on an integrated shared word-processing package (word processor for short). We discuss methodological approaches for analysing conversations where notions such as participants' perspectives and the learning context become prominent. These general considerations are derived from works from the areas of Ethnography of Communication or Discourse Analysis, and more recent works which offer new approaches on multimodal transcription (Baldry & Thibault, 2006). Also in this section, we describe the coding schema we developed for use when transcribing video screenshots.

In order to closely investigate the relationships between multimodality and language learning, more particularly with respect to writing skills, section 4 analyses two writing tasks. We explain how participants play with multimodalities in order to accomplish their tasks, at an individual or group level, which is where the expected collaboration appears. Subsequently, in section 5, constraints and patterns of use between modalities are examined by considering internal (intramodal) and external (intermodal) relationships, thanks to new tools developed by our research team (Betbeder *et al.*, 2007b). Part of the conclusion draws the reader's attention to the current tendency in CALL to overlook writing production when working in synchronous environments.

2 The research experiment: population and environment

2.1 The CoPéAs project

The ESP course was designed as part of the research project CoPéAs (*Communication Pédagogique et environnements orientés Audio-Synchrones*) run between the Université of Franche-Comté (France) and the Open University (UK) in 2005, involving sixteen French-speaking students divided into two groups of eight according to their level (false-beginner or advanced learners). Each group was consistently tutored by a different English-native tutor at the Open University proficient in designing pedagogic materials for online distance learning. Tutors and learners met in the audio-synchronous environment for eight sessions from one to one-and-a-half hours per session over a ten week period. Learners worked at home with their own computers. In addition, they used an asynchronous learning management system in order to consult instructions and publish individual pieces of written work. The course aimed at developing vocational English and competences in open distance teaching through spoken and written English (for further details on the CoPéAs project, see Chanier *et al.*, 2006). Learners were asked to actively monitor collaborative tasks, in group or subgroups, according to the assignments' instructions, whereas the tutor was in charge of the management session, moving in and out of rooms and supplying intermittent scaffolding when needed.

Our pedagogical approach is learner oriented, task-based learning oriented, and derived from socio-cultural theory (Wertsch, 1991) and situated learning (Lave, 1991). The study presented here will focus on the less proficient group (false beginners) where some learners had not practised the target language for between fifteen and thirty years.

The research protocol includes audio and video recordings in the form of screen captures of the AGSE, saving learners' individual and collaborative productions, and pre-questionnaires and post-interviews of the tutors and learners. The sample corpus thus includes 37 videos recorded over 27 hours and 512 files (productions, audiograms of the interviews, questionnaires, etc.) comprising 35Gb.

2.2 Characteristics of the AGSE

The AGSE used in our experiment is Lyceum,[1] developed and used within the Open University and designed to facilitate distance tutorials. Its structure allows tutor and learners to meet each other synchronously. The different participants connected to the environment are able to communicate orally in real time, participate in textchat, and read/modify simultaneously textual or graphic productions. The relevance of Lyceum to language learning has already been stressed by various authors, as noted above.

In Lyceum, all participants, both tutor and learners, share the same interface and the same rights. The interface is composed of three components, as shown in Figure 1:

- Spatial component (frame1): participants move from room to room or from document to document within one room. Participants can be located thanks to one grey rectangle in the spatial component, for example, here the participant is in room 101. It is also possible to see who is in the lobby. The participants are only aware of each other (audio, graphic, textchat, production) if they share the same room. They are then listed in the communication component (frame 2).

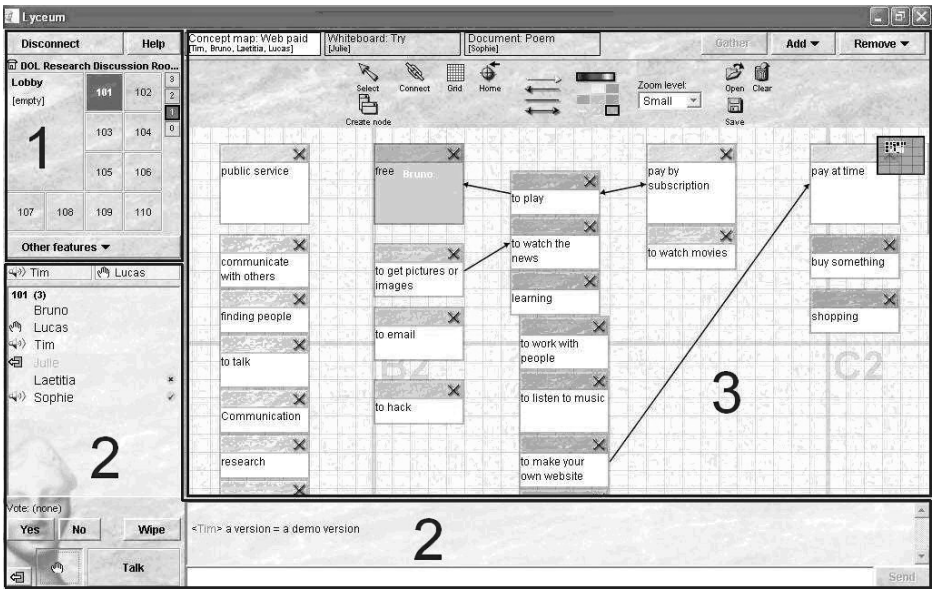

Fig. 1. The Lyceum interface and its three components

1. Lyceum: http://kmi.open.ac.uk/projects/lyceum

https://doi.org/10.1017/S0958344008000426 Published online by Cambridge University Press

- Communication component (frame 2): this includes the audio, vote and textchat tools. Each participant can, at any time, talk to the others by clicking on the button "Talk" (eg. Tim and Sophie), raise a hand to request permission to speak (eg. Lucas), vote (tick) "Yes" (eg. Sophie) or "No" (eg. Laetitia) to answer a question posed to all the participants, or make a collective decision. It is also possible to signal one's departure (eg. Julie). The textchat is another tool in this communication cluster.
- Shared editing tools component (frame 3): three kinds of shared editing tools are provided: a whiteboard which allows learners to write, draw and import images or text; a concept map for writing and organizing information; and a word processor (mistakenly labelled 'Document' in the interface) providing the opportunity for several participants to write in a single text. Up to five documents, generated by these tools, can be open at the same time. Every participant can only see and work with one document at a time, however, which means that participants who share the same room and communication tools may not view the same document. Icons at the top of the frame display which participant in working with which document. Everyone can add or delete a document, save or download changes to a document, and participate in every open document.

3 Multimodality in synchronous environments: definition and methodology of analysis

3.1 Defining multimodality in the AGSE

The AGSE supports multiple modes of communication or tools by which people construct their discourse. Moreover, all of the participants' movements and communicative acts in the different rooms and documents are easily identified. Whether the communication be textual, spoken, graphic, iconic or spatial, each mode offers many venues for sharing information (see Table 1). For example, the written linguistic mode can be realised within the different modalities of textchat, word processor and the whiteboard (on which textboxes may be created). A single mode may therefore be associated with several modalities or with a single one, as for the speech mode and the audio modality. In the following data, we evaluated two modes (written language and spoken language) and three modalities (audio, textchat, and word processor [WP]). In Lyceum, the WP is monomodal (i.e. cannot include graphs or images), but its use is embedded in a multimodal communication.

Table 1 *Correspondences between modes and modalities in Lyceum*

Modes	**Modalities**
Textual	textchat, word processor, conceptual map, whiteboard
Spoken	Audio
Graphic	Conceptual map, whiteboard
Iconic	Vote, leaving/entering the room, away for a moment, raising hand, currently speaking
Spatial	Movement (room + document)

https://doi.org/10.1017/S0958344008000426 Published online by Cambridge University Press

Participants can communicate from a large semiolinguistic repertoire of particular interest in language learning research. The richness of such a repertoire requires the organisation of all these modes. Thus, Kress and Van Leeuwen (2001: 20) define multimodality as:

> the use of several semiotic modes in the design of a semiotic product or event, together with the particular way in which these modes are combined – they may for instance reinforce each other […], fulfil complementary roles […] or be hierarchically ordered.

3.2 Methods for the analysis of language learner conversations

Since the 1990s, a number of studies, from Discourse Analysis and Conversational Analysis, deal with multimodal communication (eg., Baldry & Thibault, 2006; Scollon & Levine, 2004) and its specificity in the learning context (Kress *et al*, 2001; Lamy & Hampel, 2007). They all define multimodality as a dynamic process of meaning-making. Multimodal communication is co-constructed through interaction between participants or elements of semiotic resources, and cannot be understood piecemeal but only as a whole, albeit resembling a patchwork. As in any form of communication, the meaning intended by the speaker may differ from that perceived by the addressee. Thus, multimodality cannot be studied as a static composite production. Our aim is to define the specificity of the online virtual classroom discourse in AGSE, highlighting multimodal communication rules governing sequences, or sets of transactions (Stenström, 1994). This first level of analysis then permits us to question individual preferences and contextual preferences, in terms of multimodality.

From a methodological viewpoint, when organizing the data a first step concerns transcription. Each modality has to be identified in order to be transcribed from non-textual data to textual data. Thus, acts are defined, then tagged. The second step is to analyze the various and simultaneous modality combinations. Thus, two scales of analysis may be distinguished. On a macro scale, balances between modalities and their combinations are analysed. Here, primary consideration is given to the choice of one modality over another to express ideas, rather than the content. At a micro scale, the succession of acts in one mode/modality within a sequence is the main concern, and the content has to be analysed (see Figure 2). In this article, the focus is on the macro scale analysis.

3.2.1 The participant's perspective

A majority of researchers on multimodality derive their approaches from the Ethnography of Communication based on the studies of Malinowski, who discussed the notion of context and culture. The notion of action (i.e. what is being done in the situation with various semiotic tools, called resources) is prevalent. The focus of the analysis is not the "potential" meaning of a text,[2] but the way people interpret the text in

2. We define 'text' following Halliday (1989: 10), as "any instance of living language that is playing some part in a context of situation […]. It may be either spoken or written, or indeed in any other medium of expression that we like to think of".

https://doi.org/10.1017/S0958344008000426 Published online by Cambridge University Press

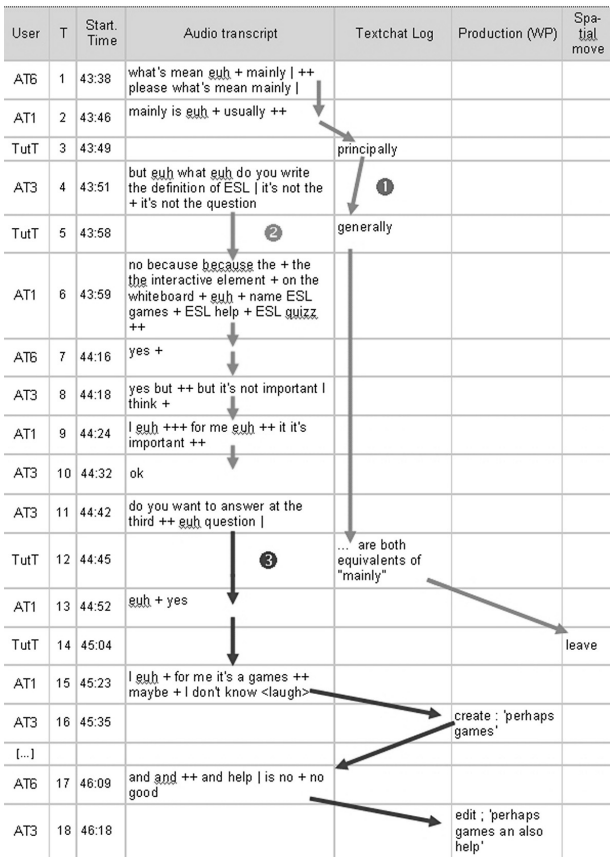

Fig. 2. Simplified versions of a tabular transcription of a screen video, extract from S5.3

different situations and activities. Participants, through their actions, demonstrate how they understand the situation, "what is going on", to quote Goffman, and what is the focus of their attention, that is, what they consider as relevant in a composite communication. Following Halliday's (1978) social theory of communication, we argue that in verbal interactions with others we have at our disposal a network of options, or sets of semiotic alternatives, which are realized through sets of options of the semantic system. The alternatives selected within the network of meanings can be considered as traces of decisions made by sign-makers (participants) about what is the most appropriate and plausible signifier for the expression of the intended meaning in a given context (Kress, 1997; Kress *et al*., 2001). Following this tradition, we chose here to analyse the learning process from the learners' and tutor's perspectives, from choices they made from a set of available meaning-making resources, in a particular situation at a given moment.

3.2.2 The notion of context

The notion of context is of the utmost importance in the study of every interaction (Goodwin & Duranti, 1992; Goffman, 1974; Sinclair & Couthard, 1975). And it is even

https://doi.org/10.1017/S0958344008000426 Published online by Cambridge University Press

more fundamental in distance interaction occurring in an AGSE. As described above, the AGSE context can be multifaceted due to the opportunities which participants have to move from one room to another or from one document to another. Following Goodwin and Duranti (1992:3), "the notion of context involves a fundamental juxtaposition of two entities: (1) the focal event; (2) a field of action within which that event is embedded". What is the focal event in an AGSE? As Jones said (2004: 27) "In the 'digital surround' created by new communication, communication is more polyfocal". However, could participants get lost among the multiple possibilities offered by this type of learning environment? Our experiment shows rather that learners make consistent individual choices to participate in multimodal discourse. It is possible to discern "focused engagements involving clear and discernable involvements" (Goffman, 1983). They also make collective choices. An AGSE can be depicted as an environment of mutual monitoring possibilities, characterised by "the moment-by-moment shifts of alignment participants bring into interaction to signal 'what they are doing' and 'who they are being'" (Goffman, 1964).

Taking into account these two dimensions, our work develops an actional perspective whereby the nature of the action influences the choice of multimodal component. The 'mode action', as described by Kress *et al*. (2001), is fundamental in any communicative situation. Therefore, the diversity of actions which may be performed requires a heterogeneous code of transcription to symbolize verbal and non-verbal modes, oral and written forms. The unit chosen to describe interactions is the "act". As Baldry and Thibault remind us (2006: xvi) "Transcription is a way of revealing both the co-deployment of semiotic resources and their dynamic unfolding in time along textually constrained and enabled pathways and trajectories".

Three dimensions are taken into account so as to "cut" the text into phases:

- The dynamism of the text (the text is considered in interaction);
- The course of the action (the text is studied historically, from a longitudinal perspective);
- The meaning-making unit (shared understanding).

Each modality is not analysed alone. Multimodality is seen as a cluster of modalities connected to each other. We do not therefore take into account any action communicated through one single modality (only via the word processor, for example). As Baldry and Thibault (2006: 19) suggest:

> Text is always multimodal, making use of, and combining, the resources of diverse semiotic systems in ways that show both generic (standardised) and text-specific (individual, even innovative) aspects.

3.2.3 Developing a coding schema

Our coding schema represents, therefore, each communicative act. Every act, whether verbal or non-verbal, is identified by a time and duration, i.e. beginning and end, and is attached to a workspace, defined here as a basic frame (space + time) in order to describe participants' actions within one collaborative tool (concept map, whiteboard or WP). The notion of workspace is important because actions occurring in one space at a

https://doi.org/10.1017/S0958344008000426 Published online by Cambridge University Press

given time may not be perceived by participants located in a separate space at that time. Each act, independent of the modality to which it belongs, is also placed in a sequence. Every Lyceum session is divided into sequences, when participants move from one space to another, when topics of conversation change, and according to the scenario. This might be: tutor's guidelines in one common room, group divided into sub-groups in separate rooms, and meeting in the common room in order to share results. An act is defined by the preceding attributes and by an actor, a modality (audio, vote, textchat, production), and its value, that is, what has been done or said (for further details, see Betbeder *et al*., 2007a). Figure 2 displays a series of acts, extracted from the fifth session, from sequence 3 (S5.3), which corresponds to one of the two writing activities discussed in this paper. Attributes such as end time and workspace have been removed in order to simplify the presentation. Silences, which we differentiated from "pauses" and analysed as such, are not represented in this example.

In this extract, three learners (AT1, AT3, AT6) are working in a sub-group. They complete the quiz (10 questions) provided by the tutor at the beginning of the session. The tutor (TutT) came into the room while AT1 was writing an ESL definition with the word processor, and using the word 'mainly' to write the definition. The extract starts when AT6 orally requests a clarification of this word (Turn T 1). The course of the conversation can be followed using the green arrows (number 1):

- AT1 replies orally (action 2)
- then the tutor answers 3 times in the textchat (3, 5 and 12)
- The learners apparently do not pay attention to what the tutor has written in the textchat. Then the tutor leaves the room (action 14).

A second audio only conversation follows the orange arrows and starts with AT3's disagreement, as s/he sees no reason for talking about ESL (action 4). The conversation ends with turn 10 when AT3 agrees. A few moments later, AT3 starts a third conversation by proposing to switch to the third question of the quiz. This conversation (blue arrows) alternates between audio and word processor modalities.

Actions occurring in the shared editing tools component are transcribed as a series of production acts ('prod', for short). In the word processor, different types of acts and a unit of action have been defined in relation to the functionalities of the tool, as shown in Table 2.

After having defined our framework, we can now qualitatively study the writing process in L2 and understand (1) how collaborative work can support the development of writing competence, and (2) the effects of the AGSE on writing competence.

Table 2 *Acts and unit of analysis in the WP*

Type of acts	**Unit taken into account**
create	Paragraph (write)
select	Paragraph
edit	Paragraph (modify form and/or content)
delete	Paragraph

https://doi.org/10.1017/S0958344008000426 Published online by Cambridge University Press

4 Analysis of two writing activities

4.1 Aims of the ESP course

The overall objective of the course was to enable learners to develop skills for working with foreign partners in the distance learning field. Students practised their English by doing some complex tasks involving reading/listening and production activities in the AGSE. The main challenge of the course was to develop the participation of false-beginner learners of ESP. Learners were also asked to do writing tasks collectively. These tasks were seen by the tutors as an opportunity to negotiate meaning and form orally.

The pedagogical aim of the writing tasks is to help learners improve their writing competence by developing their awareness of the writing process. A "learning-by-doing" approach has been chosen. Before the session, learners were given the option to take notes on vocabulary and content relating to the topic to be studied. During the session, they were asked to write a text collaboratively and from scratch, permitting them to reduce the cognitive load by focusing on a limited number of aspects at a time. As Dam *et al*. (1990) indicate, while writing, learners are faced with three main difficulties: linguistic problems, lack of automatic strategies for producing a text, and sociocultural issues. ESP task design thus targets the development of L2 writing strategies for false beginners. Therefore, the writing tasks proposed during the course are much more centred on the writing process (planning, writing, revising, correcting and publishing), rather than on the correctness of the production itself.

We report here two types of writing activity built around using the word processor: a questionnaire, based on the selection of pertinent items contained in the web-site "English-Club", where learners have to extract relevant information and reformulate it, and a guided production activity which evaluates the AGSE used during the training. The first activity is found in a subsection of session 5 (S5), located approximately in the middle of the course. The false beginners' group had been divided into two sub-groups named gr1 (N=3) and gr2 (N=3), seen respectively in sequences 2 and 3. Two learners were absent. The tutor gave instructions orally and moved between gr1's and gr2's rooms. This activity lasted fifteen minutes. The second activity lasted approximately thirty minutes, was also designed for the same sub-groups and occurred at the end of the course, in session 8 (S8). Learners reported after the experiment that they considered this activity much more difficult than the one in session 5.

4.2 When writing the 'text' with modes and modalities

The analysis presented in this section concerns the word processor and its correspondences to other modes and modalities. In this section, we examine the way participants (principally learners) combine modalities with the word processor in order to write a text collaboratively, their individual strategies, and their collaborative practices.

Table 3 offers an overview of the distribution of acts according to modalities in the two writing activities and in each sub-group.

4.2.1 Considering writing acts only

Let us start by considering only one modality, the one corresponding to writing acts (the

https://doi.org/10.1017/S0958344008000426 Published online by Cambridge University Press

Table 3 *Distribution of acts according to modalities. 'Prod' refers to production acts in the WP*

Session.Sequence	Sub-group	Distribution of acts according to modalities
S5.2	gr1	79 acts: 31textchat > 25 audio > 23 prod
S5.3	gr2	85 acts: 62 audio > 17 prod > 6 textchat
S8.2	gr1	337 acts: 199 audio > 173 prod > 65 textchat
S8.3	gr2	300 acts: 166 audio > 104 prod > 31 textchat

prod modality in the WP). As mentioned previously, we distinguish four types of acts (see Table 2): creating, editing, selecting and deleting. Figure 3 displays the proportion of acts performed in both groups in S5 and S8 and shows that the total number of writing acts is seven times higher in S8 than in S5 (see Table 3). Acts of creation show a deeper involvement of the writer than acts of editing, and even more than acts of selection or deletion. In session 5, the proportions of creating and editing acts are the same in both groups and the total number of acts almost equivalent. They represent between 70% and 80% of the total amount. Apparently both groups managed the writing of short sentences and clauses for the quiz in the same way, alternating between creating and editing.

The situation is strikingly different in session 8. Firstly, creating and editing are no longer balanced in the two groups. Secondly, each sub-group performed the activity in a very different way: in gr1 the number of editing acts is greater than the number of creative acts whereas it is the opposite in gr2. A deeper involvement in the writing process seems to have occurred in gr2.

But examining one modality at a time is like trying to understand the technique behind a monochrome painting. We need to examine 'the co-deployment of different semiotic resources and their dynamic unfolding' (Baldry & Thibault, 2006: xvi) in order to understand how learners create the picture.

4.2.2 A hierarchical organization

Figure 4 displays the proportion of acts per modality (vote, audio, textchat and WP), per

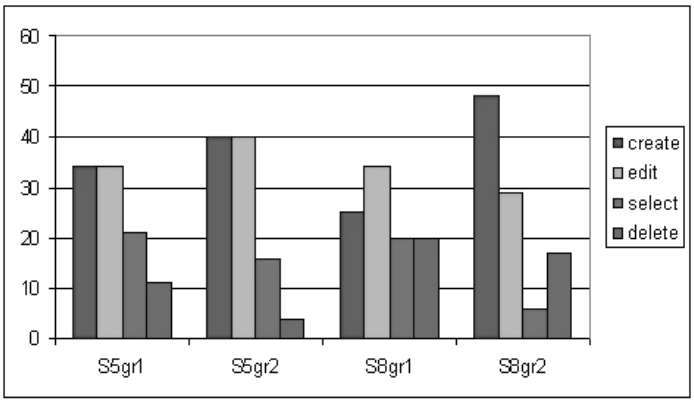

Fig.3. Percentages of type of acts in the word processor per subgroup and activity

https://doi.org/10.1017/S0958344008000426 Published online by Cambridge University Press

subgroup, performed during the two activities. It is surprising to note that, even when involved in a writing task, learners perform most of the activity orally (see, S5, gr2), by discussing what and how to write. When comparing sub-groups and activities, we can see (1) the variety of combinations performed by learners (textchat dominated in S5 within the group who had the lower L2 competence; gr1 and gr2 used the vote in S8, a more demanding activity), and (2) the prevalence of the speech mode.

These three verbal modalities (audio, textchat and WP) comprise the repertoire for language production and play complementary roles in supporting the writing process. This may be explained by examining their very natures, which differ in terms of synchronicity and duration:

```
Synchronicity: audio > textchat > WP
Duration: WP > textchat > audio
```

Audio is more 'synchronous' than textchat. Turn-taking happens more rapidly, there is no need to type, but also no possibility of rephrasing. Intervening with the word processor is a slower process. Only one participant at a time can edit a paragraph (but several paragraphs can be edited by different users). However, language has more permanence in the WP than in the textchat, and is consistently available. In its turn, production in the textchat lasts longer than oral utterances, at least before it disappears from the current window; reviewing past turns in the chat by scrolling is rare. These attributes may influence learners' choices while working: it is easier to comment quickly on a piece of text orally rather than by textchat. These elements are described in the next sections.

Comparing the use of the multimodal repertoire between S5 and S8, it appears that the use of the speech mode differs within the same group. Various factors, however, affect the use and organization of each multimodality.

4.2.3 Participant and group perspectives

Individual participants employ modalities in different ways. During the quiz activity of session 5 in sub-group 1 (gr1), learners performed 79 acts evenly distributed among

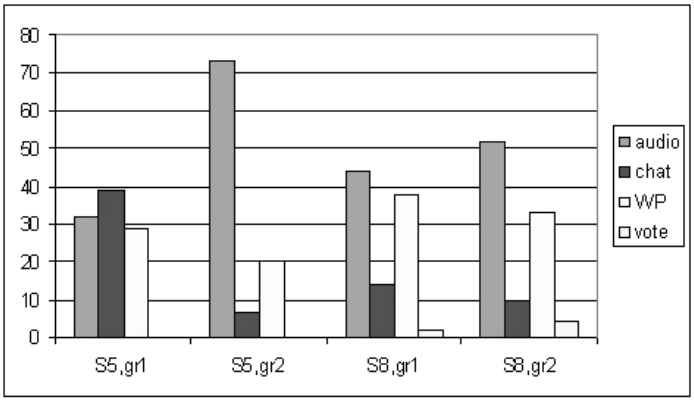

Fig. 4. Proportions of modalities used in sessions 5 and 8 per writing activity and per sub-group

https://doi.org/10.1017/S0958344008000426 Published online by Cambridge University Press

modalities (see Table 3). Figure 5 displays individual practices. Productions in the word processor are equally distributed among the three learners, and learner AT5 uses the three modalities equally. However, learners AT4 and AT2 adopt the opposite behaviour. AT4 prefers WP and audio, whereas AT2 uses textchat to complement the WP. Their use of textchat and WP are proportionally inversed.

During the same activity, sub-group gr2 performed 85 acts distributed among the three modalities in a very unbalanced way (see Table 3 and Figure 5). The speech mode is predominant; textchat is rarely used, except by the tutor; one participant took control of the word processor.

The detailed analysis of the individual practices also reveals a singular distribution of roles among the learners from one group to another. Whereas learners share the same role in gr1, all of them writing and correcting the text, learners in gr2 distribute roles and functions.

The written production process in gr1 can be characterized as 'cooperative', with each participant contributing in a different way to the achievement of a common performance. The 'modal density' (Norris, 2004) for that space-time is concentrated on the WP with some sporadic use of audio and textchat, in order to solve difficulties. The negotiation among learners comes *a posteriori*, during the correction phase of the writing process.

In gr2 only one learner (AT3) takes on the role of writer. The other two participate orally to construct the text. The use of another mode and modality (here speech and

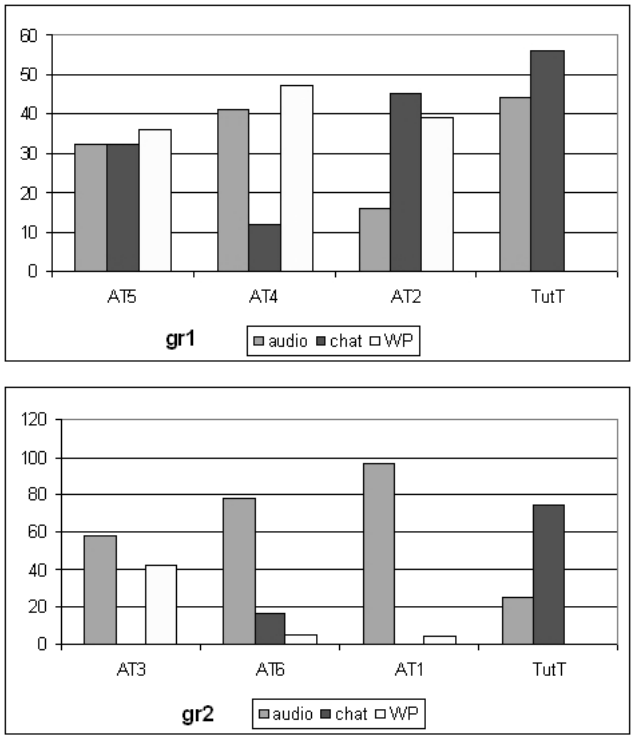

Fig.5. Percentage of modalities per participant in subgroups in the written activity of session 5

https://doi.org/10.1017/S0958344008000426 Published online by Cambridge University Press

audio) helps learners with difficulties to participate at different levels and to different degrees in the writing process. The interventions made by AT3 in the WP are the result of speech acts by the two other learners who are co-constructors of the meaning and also correctors, sometimes by text, often orally). Figure 2 illustrates this phenomenon, which occurred during the discussion tagged 3 (blue arrows). In gr2, negotiation takes place orally before writing in the WP, demonstrating a collaborative way of creating the text.

4.2.4 Contextual factors

The previous focus on the modalities per participant indicates how individuals in a given context perceive the function of each mode and modality. Nevertheless, contextual factors may have an impact on the participants' choices. Figure 6 compares subgroups during the second written task in session 8. As far as gr2 is concerned, this sub-group adopted the same collaborative model as in session 5, with AT3 remaining the group's writer. However, the situation in gr1 varies greatly from the one encountered in S5. Note that the intervention of the tutor in the WP relates only to the act of saving the document. He does not intervene in the text.

In S5, learners only had to pick sentences or words in a referenced hypertext to fill in a

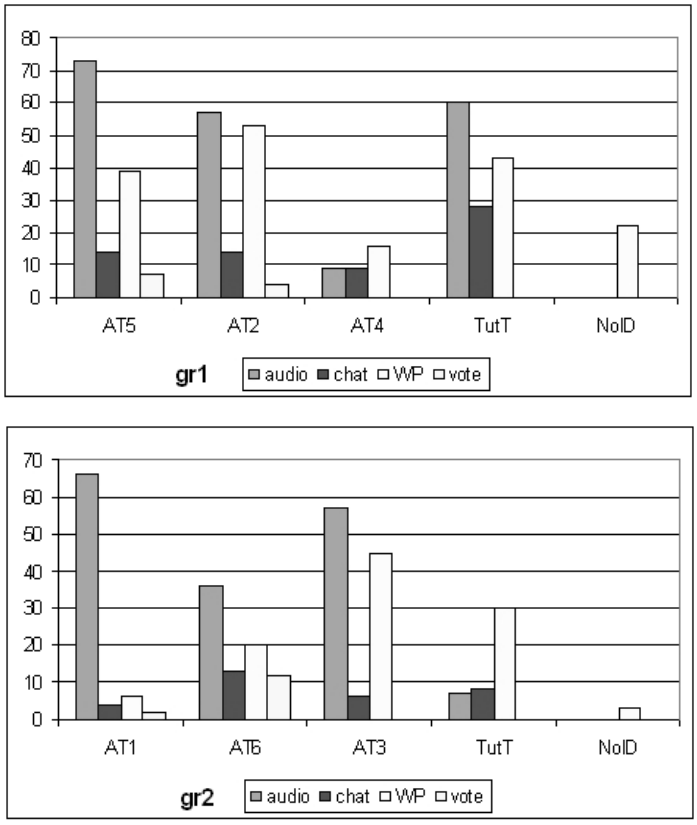

Fig.6. Percentages of modalities per participant in sub-groups in the written activity of session 8. 'NoID' refers to action in the word processor where the writer could not be identified.

https://doi.org/10.1017/S0958344008000426 Published online by Cambridge University Press

questionnaire, whereas in S8 the activity requires the learners to formulate full sentences and express opinions. Accordingly, in the latter situation, learners changed their multimodal strategies when faced with several difficulties. In gr1, the dual strategy including speech mode and WP is used by AT5 and AT2 (AT4 encountered sound difficulties). The context challenges individual preferences. For example, AT2 does not prefer using audio (this fact came out of other analyses and post-interviews). But here, the bi-modal strategy shared with AT5 is successful. She is even the one who proposed to compensate for written difficulties by dividing the writing process into different steps (conceiving the meaning orally, followed by a first formulation in the WP, then a correction orally or in the WP).

In the other sub-group, gr2, the interesting aspect is that in session 8 learners use the textchat alongside the WP in order to propose, enrich, and correct the text typed in the WP. The textchat here functions as a "writing laboratory" where everything is allowed. On the contrary, in session 5 gr2 seem to use the WP only for the final version of the text, trying to achieve the best content and form before putting it into the WP.

5 Unfolding intra and intermodal constraints in the multimodal discourse

From a discourse analysis perspective, the analysis of the multimodal organization in a given context highlights the way modes and modalities are intertwined to compose a discourse by which the collective action can be displayed. In the framework of our research project, CoPéAs, a pattern recognition tool was developed (Betbeder *et al*., 2007b), based on Manilla's Winnepi algorithm (ibid). The issue is to try to identify, among recurrent patterns, systematic ways of accomplishing actions with modalities. Patterns can be represented as a kind of rule with a left hand and a right hand side and a connector '=>' meaning 'followed by'.

```
(1)     modality1, participantX] => [modality2, participantY]
(2)    [modality1, participantX], [modality2, participantY]
       => [modality3, participantZ]
```

In (1) the algorithm considers two consecutive acts. It can be phrased as 'the algorithm found a tendency (a recurring pattern) where modality1 used in an act by participantX is followed by another act where modality2 is used by participantY". In (2), three acts are considered, i.e. the algorithm searches for a recurring sequence where two acts are followed in a systematic way by a third.

5.1 Intramodal patterns

Firstly, let us consider patterns extracted by the algorithm where modalities on the left and the right hand sides are the same, i.e. intramodal patterns. In our corpus, for all sessions running from S1 to S8, the general trend is that 75.6% of patterns are monomodal. These patterns have a form like the ones in (3) and (4) where audio acts are followed by audio acts, or in (5) and (6), where acts which are accomplished with one modality share editing tools (WP, white board or conceptual map) and are followed by acts in the same modality.

https://doi.org/10.1017/S0958344008000426 Published online by Cambridge University Press

```
(3)    [audio, x] => [audio, y]
(4)    [audio, x], [audio, y] => [audio, z]
(5)    [prod, x] => [prod, y]
(6)    [prod, x], [prod, y] => [prod, x]
```

Indeed, this intramodal organization automatically computed is typical of the corpus extracts analysed in our study. Examples (7) illustrate it within the audio modality and (8) within the word processor tool.

```
(7)
[AT3, audio]: do you want I write in the document
[AT1, audio]: yes ok XXX right++ok I (euh++)
(8)
[AT2,(WP select)]: I like when you want my opinion on such or such thing
[AT5, (WP edit)]: I like when you want my opinion on such or
    such thing that enables me to make efforts so that me am
    included/understood"
[AT4, (WP enter)]:
[AT2, (WP create)]: we like when we need to give your opinion
```

This "modal density" (Norris, 2004) shows that at some points in the action participants focus on one single modality, which becomes prevalent in their interaction in order to perform the target action. The notion of polyfocalisation is defined here as successive monomodal phases. The difference between the two sub-groups and between the two activities reveals that the multimodal organization is flexible according to what participants think is the most efficient for them.

5.2 Intermodal patterns

Secondly, if we consider intermodal patterns (left and right hand sides display different modalities), they only represent a minor part of the overall number of patterns but some of them may occur a significant number of times within a given activity, more particularly in a sub-group's work. In sessions S5 and S8, we can find recurrent bi-modal patterns (audio and production / textchat and production) such as (9) and (10), where there is a switch from audio to production (WP). In (9), one speech act accomplished by AT1 and followed by one production act by AT3 'generates' a speech act by AT6. In (10), the same actor (AT6) accomplishes one communicative goal through the successive use of two modalities, talk first, then writing in the word processor. Example (11) illustrates pattern (9).

```
(9)     [audio, AT1], [prod, AT3] =>[audio, AT6]
(10)    [audio, AT6] =>[prod, AT6]
(11)
[AT1, audio]: euh in help euh interactivity is euh ++ FAQ+
[AT3 (WP suppress), (WP edit)]: help only
[AT6, audio]: yes ++ just+choose help++ok
```

https://doi.org/10.1017/S0958344008000426 Published online by Cambridge University Press

The notion of "modal density" here takes on another significance and highlights the fact that the complexity of the action experimented with by participants generates the use of more than one modality to accomplish the target action. The intermodality appears in our corpus as scaffolding strategies to participate in and achieve the collective action (a distributed multimodality). The polyfocalisation is characterized by the integrated use of various modes and modalities which required that participants pay attention to several, oral and visual, elements of the environment. The addition of another mode has to be negotiated in order to be perceived by the participants. For example, in Figure 2, in the discussion numbered one (green arrows), the tutor gives several explanations in the textchat after a request by AT6. But learners are involved in a longer term task in which the textchat is not prevalent. They ignore the tutor and he eventually leaves the room.

5.3 Summing up the multimodal layout

The prevalent multimodal organization is a succession of monomodal phases. Yet the analysis of the work in sub-groups shows that participants use multimodal strategies more often. Therefore, "modality-switching" is a strategy frequently used by learners involved in a collaborative writing task. Two bimodal pairs have been identified and each mode has its own function:

audio (process-oriented)+ prod (product-oriented)
textchat (product-oriented) + prod (product-oriented)

From a pedagogical standpoint, it seems that intermodal strategies enhance written participation, with the writing task performed in the textchat and in the WP. The textchat complemented by the WP is used for several purposes: proposal of draft versions, content focus on particular words of the text in order to discuss their meaning, or when learners focus on the norms (correction, appropriateness) of the written text after having discussed it.

In addition, intermodal strategies support the writing process. The combination of textchat, audio and WP occurs in the corpus as follow:

- Textchat is used to improve the text, especially the form of the message (cleaning procedure).
- Audio is used to comment on and negotiate the content of the message after a first written version.
- Audio is used to provide a first version of the written message (in L2 or even sometimes in L1), and to facilitate the writing process (the deep procedure is realized through audio and WP. The focus here is on the content rather than on the form).

```
(12)
[AT3, (WP, edit)]: to have comparaison between web sites, to
know more criterias for a good site
[AT1, textchat]: according to the different criteria ?
```

https://doi.org/10.1017/S0958344008000426 Published online by Cambridge University Press

In (12) the textchat shows, in an interactive manner (note the question mark), another version of a sub-part of the written message which appeared in the word processor. Each element of the whole text, according to Halliday's perspective, nourishes the global reflection of the participants in a dynamic way.

6 Conclusions

This paper introduces an original methodological approach which aims at contributing to a better understanding of the characteristics of multimodal communication in AGSE and its relevance to the development of writing production, supported by a writing process perceived as a complex and procedural activity and as a social event. This approach may be of interest to language teachers who want to design online collaborative writing activities intertwined with synchronous communication, as well as to researchers who want to investigate further the relationships between writing and multimodal communication.

Our paper focuses on the writing process in L2 in an interactive integrated word processor and explores the way learners use the variety of modes and modalities displayed by the AGSE to create a collaborative text. The communication practices of two sub-groups of false beginners have been compared in two writing tasks of different degrees of difficulty. We based our methodological approach for analysing multimodal conversations on studies coming out of the Ethnography of Communication or Conversation Analysis, where notions such as participants' perspectives and context become prominent, to explain the communication process. We chose to analyse the learning process from the learners' and tutor's perspectives, from choices they made out of a set of available meaning-making resources, in a particular situation at a given moment. This concept of communication involves the use of a coding system, in order to be able to understand the various semiotic resources as a meaning-making unit. Interpreting multimodality means then re-building the meaning provided by the participants while communicating. Thus, multimodality cannot be studied as a static composite production. This kind of analysis needs to be extended to bigger samples and to different levels of language competence, which will be the aim of a study of the advanced group.

In focusing on the characteristics of writing in such a communication context, as opposed to other studies about multimodality in F2F where speech mode is seen as central and other modes as rather peripheral, several trends have been noted. Learners used the word processor in combination with other modalities, which highlights the strategic use of certain modes to facilitate the writing process. They also made consistent individual choices to participate in multimodal discourse, and to make collective choices. Thus, the AGSE can be depicted as an environment of mutual monitoring possibilities. The analysis of the work in sub-groups shows that participants use multimodal strategies more often than monomodal strategies. Therefore, "modality-switching" is a strategy frequently used by learners involved in a collaborative writing task. Two types of scaffolding are thus defined: collaborative scaffolding, when learners write a text together and correct each other, or when they rewrite their own production after having observed another participant's production; and multimodal scaffolding which encourages both metacognitive strategies and communicative strategies. On the

one hand, the multimodal learning environment, due to its process-oriented and collaborative nature, helps learners focus more on the writing process than on the results of their writing. The integrated word processor affords a shared visible image of the writing process. On the other hand, this environment may enrich the learner's communication repertoire (following Gumperz's perspective) by providing a new set of "compensatory strategies" which can help sustain the learner's communication.

This paper has shown the value of learners practising interactive writing in an environment offering multiple modes and modalities. Unfortunately, the current tendency in CALL is to overlook the writing process when working in a synchronous environment. In general, the focus lies on the use of video to enhance oral communication, which leads to the study of other dimensions of multimodality, such as kinesics and mimics. In our study, where speech is related to various forms of writing, the use of video would have distracted the learners. Certainly, further experiments and research are needed in the area of writing in an online multimodal environment in order to more fully understand the role of multimodal verbal communication in collaborative tasks, specifically in terms of interaction analysis and content analysis.

Acknowledgements

We would like to thank Marie-Laure Betbeder, Françoise Greffier, Muriel Noras, Christophe Reffay and Anna Vetter, from the Université de Franche-Comté, who have been involved in transcribing and organizing the data and have developed tools. We would like to express our thanks to Bonnie Youngs (Carnegie Mellon University) for her valuable help, and to the anonymous reviewers for their helpful comments.

References

Andrews, T. (1993) Videoconferencing: an interactive communication tool for distance learners. *ONCALL*, **8**(1) http://www.cltr.uq.edu.au/oncall/andrews81.html

Baldry, A. and Thibault, P. (2006) *Multimodal Transcription and Text Analysis, a multimedia toolkit and coursebook with associated on-line cours*e. London: Equinox.

Betbeder, M.-L, Ciekanski, M., Greffier, F., Reffay, C. and Chanier, T. (2007) Interactions multimodales synchrones issues de formations en ligne: problématiques, méthodologie et analyses. In : Lamy, M-N, Mangenot, F. and Niessen, E. (eds.) EPAL 2007, Grenoble, France, 7-9 juin 2007 http://w3.u-grenoble3.fr/epal/pdf/betbeder.pdf

Betbeder, M.-L., Tissot, R. and Reffay, C. (2007) Recherche de patterns dans un corpus d'actions multimodales. *EIAH Conference,* Lausanne, Suisse, 27-29 juin 2007. http://edutice.archives-ouvertes.fr/edutice-00158881

Chanier, T., Vetter, A., Betbeder, M.-L., Reffay, C. (2006) Retrouver le chemin de la parole en environnement audio-graphique synchrone. In : Dejean-Thircuir, C. and Mangenot, F. (eds.) Les échanges en ligne dans l'apprentissage et la formation. *Le Français dans le Monde*, Recherche et applications, 40: 139-150 http://edutice.archives-ouvertes.fr/edutice-00084388

Chanier, T. and Vetter, A. (2006) Multimodalité et expression en langue étrangère dans une plate-forme audio-synchrone. *Alsic*, **9**: 61-101 http://alsic.u-strasbg.fr/v09/chanier/alsic_v09_08-rec3.pdf

Dam, L., Legenhausen, L. and Wolff, D. (1990) Text production in the foreign language classroom and the word processor. *System*, **18** (3): 325-334.

Dejean, C. and Mangenot, F. (forthcoming) Gestion à deux d'une tâche informatisée d'écriture à contraintes. In: Bilous, D. and Oriol-Boyer, C. (eds.) *Actes du colloque 'L'écriture à contraintes'*. Université Stendhal-Grenoble 3, May 2000.

Goffman, E. (1964) The neglected situation. *American Anthropologist*, **66**:133-136.

Goffman, E. (1974) *Frame analysis: an essay on the organization of experience*. New York: Harper & Row.

Goffman, E. (1983) The interaction order. American *Sociological Review*, **48** (1): 1-17.

Goodwin, C. and Duranti, A. (1992) Rethinking Context: An introduction. In: Goodwin, C. and Duranti, A (eds) *Rethinking Context: Language as an interactive phenomenon*. Cambridge: CUP: 191-227.

Halliday, M.A.K. (1978) *Language as Social semiotic: The social interpretation of language and meaning*. London: Arnold.

Halliday, M.A.K. (1989) Part A. In: Halliday, M.A.K. and Hasan, R. (eds.) *Language, Context and Text: Aspects of language in a social-semiotic perspective*. Oxford: Oxford University Press, 55-79.

Hampel R. (2006) Rethinking task design for the digital age: A framework for language teaching and learning in a synchronous online environment. *ReCALL*, **18** (1):105-121.

Hampel, R. and Baber, E. (2003) Using internet-based audio-graphic and video conferencing for language teaching and learning. In: Felix, U. (ed.) *Language Learning Online: Towards Best practice*. Lisse: Swets & Zeitlinger, 171-191.

Hampel, R. and Hauck, M. (2004) Towards an effective use of audio conferencing in distance language courses. *Language Learning and Technology*, **8** (1): 66-82. http://llt.msu.edu/vol8num1/pdf/hampel.pdf

Jeannot, L., Vetter, A. and Chanier, T. (2006) Repérage des stratégies des apprenants et du tuteur dans un environnement audio-graphique synchrone. In : Dejean-Thircuir, C. and Mangenot, F (eds.) Les échanges en ligne dans l'apprentissage et la formation'. *Le Français dans le Monde, Recherche et applications,* **40**: 151-161. http://edutice.archives-ouvertes.fr/edutice-00087741

Jones, R. (2004) The problem of Context in Computer-Mediated Communication. In: Levine, P. and Scollon, R. (eds.) *Discourse & Technology, multimodal discourse analysis*. Washington, Georgetown University Press, 20-33.

Jepson, K. (2005) Conversations – and negotiated interaction- in text and voice chat rooms. *Language Learning & Technology*, **9** (3): 79-98. http://llt.msu.edu/vol9num3/pdf/jepson.pdf

Kress, G. (1997) *Before Writing: rethinking the paths to literacy*. London: Routledge.

Kress, G., Jewitt, C., Ogborn, J. and Charalampos, T. (2001) *Multimodal Teaching and Learning, the rhetorics of the science Classroom*. London: Continuum.

Kress, G. and Van Leeuwen, T. (2001) *Multimodal Discourse: the modes and media of contempory communication*. London: Arnold.

Lamy, M-N. (2006) Conversations multimodales: l'enseignement-apprentissage à l'heure des écrans partagés. *Revue Le Français Dans le Monde, Recherche et Applications*, **40**:129-138.

Lamy, M-N. and Hampel, R. (2007) *Online Communication in Language Learning and Teaching*. Palgrave Macmillan.

Lave, J. (1991) Situating Learning in Communities of Practice. In: Resnick, L.B., Levine, J.M. and Teasley, S.D. (eds.) *Perspectives on Socially Shared Cognition*. Washington, DC: American Psychological Association, 63-82.

McAndrew, P, Foubister, S.P. and Mayes, T. (1996) Videoconferencing in a language learning application. *Interacting with Computers*, **8**(2): 207-217.

Norris, S. (2004) Multimodal Discourse Analysis: A Conceptual Framework. In: Levine, P. and Scollon, R. (eds.) *Discourse & Technology, multimodal discourse analysis*. Washington: Georgetown University Press, 101-115.

Scollon, R. and Levine, P. (2004) Multimodal Discourse Analysis as the Confluence of Discourse

and Technology. In: Scollon, R. and Levine, P. (eds.) *Discourse & technology, multimodal discourse analysis*. Washington: GUP, 1-6.

Sinclair, J. and Couthard, R-M. (1975) *Towards an Analysis of Discourse. The English Used by teachers and Pupils*. Oxford: Oxford University Press.

Stenström, A. B. (1994) *An introduction to spoken interaction*. London: Longman.

Stevens, V. and Altun, A. (2002) The Webheads community of language learners online. In: Syed, Z. (ed.) *The process of language learning: An EFL perspective*. Abu Dhabi: The Military Language Institute, 285-318.

Stockwell, G. (2007) A review of technology choice for teaching language skills and areas in the CALL literature. *ReCALL*, **19** (2): 105-120.

Vetter, A. (2004) Les spécificités du tutorat à distance à l'Open University. *ALSIC,* 7: 107-129. http://alsic.u-strasbg.fr/v7/vetter/alsic_v07_06-pra2.htm

Vetter, A. and Chanier, T. (2006) Supporting oral production for professional purpose, in synchronous communication with heterogeneous learners. *ReCALL*, **18** (1): 5-23.

Wang, Y. (2004) Internet-based desktop videoconferencing in supporting synchronous distance language learning. *Language Learning and Technology*, **8**(3): 90-121. http://llt.msu.edu/vol8num3/pdf/wang.pdf

Wertsch, J.V. (1991) A Sociocultural Approach to Socially Shared Cognition. In: Resnick, L.B., Levine, J.M. and Teasley, S.D. (eds.) *Perspectives on Socially Shared Cognition*. Washington, DC: American Psychological Association, 85-100.