Researching Technology-mediated Multimodal Interaction

Thierry Chanier, Marie-Noëlle Lamy

Chanier, T. & Lamy, M-N. (2017). Researching Technology- mediated Multimodal Interaction. In *The Handbook of Technology and Second Language Teaching and Learning*, Chapelle, C.A. & Sauro, S. (eds). John Wiley & Sons. Pp 428-443

Researching Technology- mediated Multimodal Interaction

THIERRY CHANIER AND MARIE-NOËLLE LAMY

Language learners have access to a wide range of tools for communication and interaction through networks, where multiple options exist for creating meaning. This chapter introduces the study of such multimodal meaning-making for language learning. The study of meaning-making through the use of technology for mediation has been pursued since the technology became widespread. An overview of multimodality in interaction is offered in the introduction to a Special Issue of Semiotica by Stivers and Sidnell (2005); for multimodality in communication, see Kress (2010); for multimodality in classroom-based language learning, see Royce (2007) and Budach (2013); for a discussion of digital mediation see Jones and Hafner (2012), and for a summary about technology-mediated language learning see Hampel (2012). These sources provide a starting point for the investigation of multimodality in language learning, which is relatively recent computer-assisted language learning (CALL) research. In view of the importance of such meaning-making resources for learning, this chapter introduces the meaning of "mode" and "multimodality" before introducing the research and theoretical issues in the study of multimodality in language learning. It then describes a corpus-based approach for studying multimodality in CALL, drawing attention to important methodological factors such as transcription that come into play in conducting such research.

Multimodality in online learning

During online communication in a second language (L2), learners orchestrate various resources including language, in its written and spoken forms, as well as images, colors, movements, and sounds. Responding to even a simple written post requires at minimum two such resources: the linguistic mode, that is, the written language, and the visual mode, which involves the choice of fonts and the organization of spaces on the screen. In addition to language and image, online tools such as a floating magnifier facilitates the understanding not only of written communication but also of pictures. In online shops such tools can be used to reveal enlargeable images of the product to allow the shopper to see products in greater detail. The floating magnifier tool in itself has no meaning but the enlarged image or written text probably does. For language learners, the floating magnifier may be used to support a vocabulary or grammatical

gloss, and an image may carry a cultural reference. The physical material also includes keys to be pressed for text creation, screens to be tapped, pads to be stroked or hotspots to be clicked for opening up video or audio channels. In computer-mediated interactive language learning (henceforth CMILL), learning is affected by the resources that are available to learners and their use. Therefore, the design of learning activities and research on their use needs to take into account of the materiality of the modes available to learners and how they are used to create meaning multimodally (Lamy 2012a).

What are modes?

Mode in linguistics refers to the resources used to express meaning. For example, Kress and van Leeuwen (2001) showed that readers make sense out of a page of a newspaper by combining their understanding of the linguistic content (linguistic mode), with their interpretation of the layout and photos or cartoons (visual mode). In other words, language users combine semiotic resources to convey messages through simultaneous realization of linguistic and non-linguistic modes in printed media. In CMILL, the resources to be co- orchestrated by participants are made up of a greater variety of modes than those in print media or audio-video alone. CMILL is carried out through the use of modes, which are accessed and manipulated with tools to carry out certain learning objectives. The integration of these three aspects of communication makes up modality. The relationships between the three are illustrated in Table 28.1.

Table 28.1 Modality as a set of relationships among objectives, tools, and modes in CMILL. Adapted from Lamy (2012b).

Main CMILL objectives facilitated	*Tools*	*Modes*
Information seeking (preliminary to engagement with tasks or people)	Screens	Dual modes: linguistic (written) and visual
Accessing and interacting with materials and people	Screens (conventional, tactile), pads, microphones, speakers	Multiple modes: linguistic (written, sometimes spoken), visual, kinetic, aural if music is involved)
Reflective activities (revisiting tutorials and conversations), sharing audio-visual material	Recording and screen capture tools	Multiple modes: linguistic (written and spoken)
Communication and interaction (e.g., Like button); create tele-presence; help with turn-taking (e.g., raised-hand button)	Hot buttons	Multiple modes: linguistic (written and spoken), visual, kinetic
Written exchanges; peer collaboration; feedback; commenting; group bonding	Asynchronous sending/ receiving channels; synchronous messaging channels	Mainly written linguistic mode, with elements of visual mode
Oral communication and interaction; collaborative work; feedback; group bonding	Webconferencing	Multiple modes: linguistic (spoken and some written), visual, kinetic

The objectives refer to the types of communication and learning activities that the students engage in such as information seeking as part of a communication task, reflection on their past work, and peer collaboration to produce a written product. The tools are the software and hardware configured in a way that allows learners to use them to accomplish learning objectives. These tools are socially-shaped and culturally-constructed to provide access to the modes of communication. Kress (2010) explains the connection between tools and modes, by noting, for example, that an image shows the world and a written text tells the world. Therefore modes offer "distinctive ways of representing the world" (96). So while a tool may borrow a representation from the world (a cursor may be materialized on screen as a hand or a pair of scissors) the tool merely indicates by this that its function is to grab or to cut. Its function is not to represent hands or scissors to the user. Materialized through the use of tools, modes combine together to facilitate meaning-making in learning. Thus we can

define modality as the relationship between tools and modes.

What is multimodality?

Multimodality is the complex relationship that develops between multiple tools and modes when they are co-deployed in different combinations, in learning situations to work toward particular objectives. In online audio-visual environments the complex meaning-making (or semiotic) possibilities that open up to language learners are materialized through hardware and software. New meanings emerge in learning situations through learners' physical relationship to tools (sometimes called embodiment), through participants' body language on screen (another form of embodiment), through learners' engagement with still and moving images, with sounds, and with each other's language outputs. All of this is experienced in simultaneous integration or, as some multimodality researchers put it, co-orchestration.

Another way of expressing this integration is by examining the notion of mediation. Mediation is always present in any kind of human interaction, including multimodal ones. The separate meaning-making resources of mediation in online learning are illustrated in Figure 28.1 as three circles, labeled A, B and C. Circle A represents the resources available through the participants' physical bodies. Circle B represents the resources that are available through the technology. Circle C symbolizes the meaning-making resources available through language.

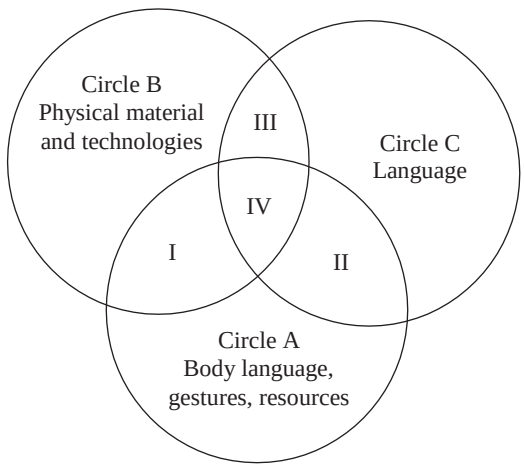

Figure 28.1 Schematic diagram showing components Mediation in CMILL.

In any situation of meaning-making, these separate components of mediation form intersecting areas (shown in the figure with Roman numerals). In offline meaning-making, only two types of resources are present: for instance when we hail friends across a coffee-bar, we are making meaning with our smiling face, our waving arm, and our cheerful call of their name (represented by Area II, where Circles C and A

intersect, that is, where language and body language meet). Language II often part of meaning-making, but Area I, at the intersection of Circles A and B, represents instances when language is not involved. For example, a piano tuner using auditory senses and hands (Circle A) to tune a piano (Circle B) would create meaning without language. Finally, an example of Area III might be two computers programmed to communicate with each other.

In online multimodality, the three components intersect: for example when we edit our photos and upload them to a social network site, we are working mainly with our hands and eyes (resources in Circle A), with our tablet/phone/laptop (resources in Circle B), with the language of our editing suite's instructions, and with our own language (resources in Circle C) to create captions and label albums. In addition if we interact with others (for example by engaging with them in "commenting" on the photographic network) we again bring in resources from Circle C. Our online interaction is thus mediated through technology, through our own body and through language (a triple overlap represented by Area IV in the middle of the figure). Area IV is therefore the space where phenomena of interest to researchers in multimodality within CMILL reside.

Research on multimodality in online learning

Research on multimodality examines two questions: "What aspect of the learning is mediated through the technology?" and "How is the learning situation experienced, especially in terms of affect?" Erben (1999) showed participants creating new discursive practices by altering the content of their exchanges (through "reducing" and "amplifying" meanings) as they strove to find workaround responses to what they experienced as an unwieldy communication platform. His findings made it clear that the "what" of communication was affected, as participants adapted to the crudeness of the mediation tools at their disposal at the time. On the other hand, Goodfellow et al. (1996) as well as McAndrew, Foubister, and Mayes (1996) reported on the "how," showing that video-based learning inter- actions were experienced as stressful, although at the same time video was seen by learners as motivational.

These two lines of investigation have continued to structure much CMILL research in recent years. First, the issue of the "what" needs to be thoroughly understood if communication distortion/breakdown is to be avoided and also, more positively, if the communicative affordances of the multimodal environment are to be maximized. In the decades that followed the pioneering studies above, scholars continued to explore many aspects of mediation as reflected in several of the articles on multimodal learning in *ReCALL 25* (2013) and elsewhere. For example, Codreanu and Celik (2013) found that tool management influenced the content of the learning while Satar (2013) analyzed the mediating effects of gaze, and Wigham and Chanier (2013a, 2013b) studied the mediation of the learning through digital gestures in *Second Life*. Dooly and Sadler (2013) looked at how *Second Life* mediated the transfer of knowledge from theory to practice, and Monteiro (2014) studied video-mediated corrective feedback. As

Wang and Chen pointed out in 2012, "in-depth research is needed to establish the extent to which visual cues mediated through videoconferencing tools are important to collaborative language learning. Such research is more urgently needed now than it was 5 years ago as broadband technology has already made good quality of video transmission a reality" (325). A wider discussion of what tools and affordances best mediate learning in CMILL can be found in Hampel (2014) and in the domain of synthetic worlds, a review of mediational tools appears in Schwienhorst (2002). Finally, Reinhardt and Sykes' (2014) edited issue on game-informed L2 teaching and learning.

The second line of research, centering on "how" the learning situation is experienced, that is, on affect, has produced literature on psychological variables. These range across anxiety, anger, regret, desire and poor self-esteem. An early study of Web use (Yang 2001) identified anxiety as a factor connected to cognitive overload. However, other anxiogenic factors in language interaction on platforms did not come to the fore until the late 2000s, with work by Hurd (2007), de los Arcos (2009), de los Arcos et al. (2009), Coryell and Clark (2009), Develotte, Guichon, and Vincent's (2010) analysis of gaze in desktop conferencing from a socio-affective viewpoint and Tayebinik's and Puteh's (2012) literature review article on self-esteem.

Scholars whose work has included both what and how of learning during the last decade are Wang and to a lesser extent her co-authors Chen and Levy. They studied desktop videoconferencing, which they later called "synchronous cyber face-to-face" from several angles: participants' perceptions of the benefits of tools (Wang 2004), focus on form in task completion (Wang 2006), task design (Wang 2007), teacher training (Wang et al. 2010), collaboration (Wang and Chen 2012) and the question of how principles of instructed language learning can be applied in these environments (Chen and Wang 2008). They found that potential learning benefits could arise in multimodal environments provided tools were used in a way that was balanced and appropriate to pedagogies, to technologies, and to audiences. They also found potential learning losses due to synchrony nerves and anxiety (including among teachers). There are indications that the affect-oriented strand of multi-modality research in CMILL is a continuing concern of scholars.

Theoretical underpinnings

Few scholars of multimodality in language learning ground their work theoretically except to claim an affiliation to "sociocultural approaches." Among those who identify a theoretical basis for their work, the literature is broadly divided into those using technoliteracy frame- works and those relying on semiotic theories. In the former category, the aim is to investigate obstacles and facilitators in the development of learner competence with the platforms and tools, while the latter are more interested in understanding how learners orchestrate meanings mediated to them through a variety of modes in the online situation.

On the one hand the technoliteracy-oriented CMILL community has long insisted that task design is key and should closely match the communicative affordances of the

environments. The practical consequences of this are outlined by Hampel (2014) "Tasks need to be appropriate to the environment, and it is crucial that activities that make use of digital environments take account of their functionalities and affordances" (18). Others have also stressed the centrality of pedagogy and task design, with studies focusing on audiographic conferencing; webcam-assisted conferencing; *Second Life*; and processing overload. For an overview on multimodality research and literacy in CMILL, see Ho, Nelson, and Müller- Wittig (2011).

On the other hand, CMILL scholars using social-semiotic theories have responded to a different priority: the need to understand a learner's meaning-making in multimodal situations. One seminal text explaining how social semiotics can account for multimodal communication is Kress (2010). However critics have complained about what they see as the over-prominent role played by language in Systemic Functional Linguistics, a theory that Kress and others (O'Halloran and Smith 2011) consider to be core to the shaping of social-semiotic theories. These critics have argued that for online language-learning set- tings, it is necessary to further root social-semiotic analysis in notions of place and embodiment. Among these scholars some concentrate on the relationship of the body with the physical environment of the virtual experience (Jones 2004; Lamy 2012a), others on silence and gaze (Stickler et al. 2007; Develotte et al. 2010), and others yet on social presence (Dahlberg and Bagga-Gupta 2011, 2013; Lamy and Flewitt 2011). However an unresolved issue in the social-semiotic analysis of CMILL exchanges remains that of their linguistic component, and of how to transcend language-based methods such as discourse analysis (Sindoni 2013) or conversational analysis, so as to fully recognize their multimodal dynamics.

Given that technology has now opened up possibilities for fully-documented, accurate, and exhaustive capture of multimodal exchanges, both the technoliteracy and the social semiotics research communities should be able to establish a synergistic relationship bet- ween their theories and the empirical data that they collect. However the complexity of multimodal environments, and the sophistication of the tools that can capture it, combine to create massive datasets and not all researchers need all data. So the choice of an appropriate window of observation for each particular project and the selection of relevant categories of data for collection, storage, and analysis are key.

The need to analyze multimodal data in education

The difficulties of collecting and handling online multimodal data have been problematized in non-educational fields. See for example Smith et al. (2011). Also, for an overview of almost two decades of semiotics-based investigation of multimodal data, see O'Halloran and Smith (2011) and O'Halloran et al. (2012). In contrast, the CMILL literature provides few methodological publications focusing on the treatment of multimodal data in language-learning con- texts online. An article entitled "What are multimodal data and transcription?" applied to education more generally (Flewitt et

al. 2009) pointed to this gap in the field, and will serve as our introduction the problems of working with multimodal data. "Drawing on findings from across a number of disciplines and research fields including applied linguistics, visual ethnography, symbolic representation and computer-mediated communication" (53), the authors outline issues such as the transcription, description, and analysis of multimodal texts and, before all else, the definition of a unit of analysis for research. They state that in dynamic texts (i.e., conversations) "units of transcription are usually measured as turns of speech, but it is questionable whether this convention is useful for multimodal analysis" (47). Because other modes come into play, Flewitt et al. suggest either linking measurement to the visual mode and using visual frames as units or timeframes can provide definitions for the unit for analysis.

However, kinesics data (how the on-screen moves of the artifacts are understood), proxemics (how distant participants "feel" they are to one another), or postural, gestural, and gaze elements should also be included. The authors conclude that "the representation of the complex simultaneity of different modes and their different structure and materiality— has not been resolved in transcription, nor have satisfactory ways as yet been found to combine the spatial, the visual and the temporal within one system" (47). Researchers can negotiate these difficulties by making pragmatic decisions about priorities for recording/ transcribing, depending on the object of the research, but in any case they need to be aware that fully represented multimodal transcriptions may be "illegible," which out- weighs the advantage of their descriptive accuracy. Finally, Flewitt et al. discuss examples from various researchers prioritizing different semiotic components but they again conclude that all of these representations "pose significant disadvantages for research dissemination, where the dominant media are still printed, written or visual formats" (52). This work has important implications for research-oriented corpus-building (see Dooly and Hauck 2012).

Methodology for developing CMILL research: The need to investigate multimodal Interactions

In online language learning means that studies should examine situations where participants (learners, teachers, natives, etc.) are involved in activities which span over several weeks, including several hours a week. The coverage of the data collected for analysis is a key factor. This is one of the reasons for adopting a corpus-based approach for systematically gathering, transcribing and coding large amounts of longitudinal multimodal data. Our corpus-based approach is intended to overcome the limitations of the approach to research introduced by Flewitt et al. (2009), which introduces different, incompatible ways of collecting, organizing, and analyzing data. Our corpus-based methodological approach which navigates from transcription to coding and analysis suggests a compatible way to support multiple analyses in addition to sharing data among researchers.

Motivations for a corpus-based approach

This corpus-based approach seeks to address a range of scientific criteria that apply to research on second language acquisition (Mackey and Gass 2005) such as validity (Do the data allow for meaningful interpretations? Is the scope of relevance of the results clear not only to our sample but to other contexts?), and reliability (Is the manner of data collection, organization, and analysis sufficiently consistent to allow for comparable results to be obtained by other researchers?). We developed a corpus-based approach based on our experience initiated in 2005 with the concept of LEarning and TEaching Corpus (LETEC) in online multimodal learning situations. It combines work on speech corpora related to first language acquisition and computer-mediated communication (CMC) corpora and its model, which encompasses multimodal interactions. It addresses issues of meaningfulness and consistency by fitting with the expectations and conceptual frameworks of researchers in applied linguistics in addition to encompassing quality criteria for gathering and analyzing data.

The experience of the language acquisition community

Researchers investigating interaction through spoken language have needs for data that are similar to those of CMILL researchers. For example, Jenks (2011, 71), who specializes in the topic of transcribing speech and interaction, reminds us this overlap:

> In face-to-face another types of multimodal interaction, non-verbal conduct (e.g. gaze, body, posture, pointing and nodding) is equally as important, prevalent, and multifunctional, as stress, intonation, and voice amplitude.

In language acquisition, researchers have to start by defining what kinds of observations and measures will best help capture the development process over time. Generally interactions need to be captured in authentic contexts rather than in laboratory conditions. For example, decisions have to be made, when studying discourse addressed by adults to children (i.e., Child Directed Speech), about the number of subjects to be studied in order to fulfill scientific criteria, about the type of situations in natural settings, the length of every window of observation (What time of the day? What child activity? Should the researcher be present or absent?). Choices need to be made concerning the repetition of the observations. Hoff (2012) provides a good introduction to these issues. Opportunities and pitfalls have been extensively studied over time by the language acquisition community. All these research protocols may illuminate the way we can develop multimodal CALL research in informal settings.

The longstanding tradition of building research on corpora illustrates the benefits CMILL may expect when following a similar route. Large repositories of corpora have been built over time by an international set of researchers following a unique methodology such as CHILDES on first language acquisition (MacWhinney 2000).

The research planning is presented as a three-step process (O'Donnell cited in Segouat et al. 2010), which we synthesize as: (1) May I find appropriate data in existing corpora, or when extracting parts of different corpora? (2) What if I recombined and rearranged them with a different perspective in mind? (3) I will consider developing a new corpus if (and only if) answers to the two previous questions are negative.

This perception of research as a cumulative process, and of the analysis as a cyclical one, with researchers reconsidering previous data, mixing them with new data, measuring things differently in accordance with new theoretical frameworks, is a reality in language acquisition studies, and more generally in spoken corpora. A good illustration is presented in (Liégeois 2014) around the *Phonologie du Français Contemporain* (PFC) corpus which gathered together a community of international researchers over 15 years to pave the way for fundamental discussions about the nature of language acquisition.

Quality criteria for CMILL corpora

Since research questions in CMILL are always connected to learning situations, data collection will not only refer to multimodal interactions, listed in the previous section. It will also encompass: (1) the learning situation, (2) the research protocol, and (3) the permissions for access. The learning situation refers to the learning design if it is in a formal situation, and to other elements of the context if it is in an informal situation. It also refers to all the necessary information about the participants (e.g., learners and teachers) bio- graphical and sociological information, level of expertise not only with respect to language but also with the technological environment, and so on. The research protocol refers to questionnaires, participants' interviews, methodology for data collection and coding, and so on. The permissions for access specifies how data have been collected (how the question of ethics and rights have been taken care of? It provides the consent form used, and explains the anonymization process on raw data), and how the corpus contents can be freely used by other researchers.

In order to become a scientific object of investigation a corpus has to meet several quality criteria: systematic collection (Were the data collected systematically in view of the research phenomena at stake? Is the coverage representative?); coherent data organization: coherence for packaging the various parts of the corpus, interlinking them (e.g., video files with transcriptions), coherence when coding and transcribing; data longevity, including a short-term window in order to be able to use several types of analysis tools and collect data in nonproprietary formats; and a middle-term window in order to deposit the data, share them, and store them in an archive; human readability, including information allowing researchers who did not participate in the experiment to work with the corpus; machine readability for data storage and, beforehand, for analysis purposes; the so-called OpenData criteria, related not only to the aforementioned permissions, but also to the guarantee of continuing access to the internet and to efficient identification of the corpus on Web search, as

described by Chanier and Wigham (2016).

These criteria aim at the issues of validity and reliability by striving to make clear what the corpus represents and how usable it is for the purpose of analysis. Representativeness relates not only to the systematic way of collecting data, but also to the way its scope has been delineated (Baude et al. 2006). All these kinds of data build up what we called a LETEC corpus (Chanier and Wigham, 2016; Reffay, Betbeder, and Chanier 2012; see the Mulce corpora repository, 2013, from which 50 corpora can be downloaded). The LETEC contains not only data, but also their detailed metadata which describe: conditions of data collection, aggregation, organization, coding, general information about the learning situation, about the technological environment, and so on. These elements of information provide a basis for a meaningful analysis.

Conceptualization of multimodal acts

A critical aspect of making research interpretable across different projects is to have a scheme for classifying and analyzing the many different types of multimodal data. The analytic scheme needs to capture certain types of relevant actions performed and/or perceived by CMILL participants within a given space. Beißwenger et al. (2014) call the site of interactions the "interaction space" (henceforth IS), which is an abstract concept temporally situated at the point when interactions occur online. The IS is defined by the properties of the set of environments used by the participants. Participants in the same IS can interact (but do not necessarily do so, cf. lurkers). They interact through input devices mainly producing visual or oral signals. Hence when participants cannot hear or see other participants' actions, they are not in the same IS. Within a variety of different ISs, multiple types of multimodal acts have been studied in various research projects, from data collected and organized in LETEC corpora. All such acts can be classified as either verbal or non-verbal acts.

Verbal acts, such as those studied by corpus linguistics, are based on textual and oral modes. In synthetic worlds such as *Second Life*, every participant can choose to use audio and textchat in order to publicly communicate with other participants located close to her/ his avatar. S/he can also decide to communicate with people not co-present (Wigham and Chanier 2013a, §2.1). For the sake of simplicity, we will assume that audio and textchat can be heard or read by participants in the same location. When studying interactions between various kinds of acts, it is useful to distinguish between a verbal act which is realized as an *en bloc* message and an adaptive one. Once a textchat message has been sent to a server, it appears to the other participants as a non-modifiable piece of language (it becomes a chat turn and has lost any indication of the way it had been planned by its author before being sent). On the contrary, a participant's utterance (e.g., in an audio chat act) can be planned, then modified in the heat of the interaction while taking into account other acts occurring in other modalities of communication (Wigham and Chanier 2013b).

As regards non-verbal acts, a great deal of attention is paid in social semiotics and in CALL research to acts related to the body, whether generated by actual human body

(mediated through webcams) or by avatars in synthetic worlds. Wigham and Chanier (2013a) presented a classification of such acts (for another viewpoint see Shih 2014): proxemics, kinesics (which includes gaze, posture, and gestures), and appearance. Another type of non-verbal derives from actions in groupware. These share-edited tools, such as word processor, concept map, whiteboard can be integrated within wider environments for example, audiographic, or used besides video conference environments. They have been largely developed and studied by the computer-supported collaborative learning (CSCL) community (Dyke et al. 2011). Within these environments knowledge is collaboratively built and negotiated with interaction switching between non-verbal acts and verbal acts (Ciekanski and Chanier 2008).

Lastly we will consider non-verbal acts based on the use of icons. A first type of iconic tool is specifically oriented toward conversation management: such tools ease the turn-taking (icon raised hand), reduce overlap, restrain talkative participants, so as to encourage the more reserved ones to take the floor; they allow quick decisions and clarifications to be made (voting yes/no); they focus on technical aspects without interrupting the conversation flow (icons talk-louder, talk-more softly) or they mediate social support (clapping icon). The second type of iconic tool displays signs of participants' presence, for example icons show when a participant enters or leaves the interaction space, or is momentarily absent.

From transcription and coding to analysis

The scheme for classifying multimodal acts in CMILL situations provides a conceptual framework for categorizing multimodal data in CMILL, but to carry out research, the data also need to be transcribed so they can be analyzed. A transcription is a textual version of material originally available in a non-textual (or non-exclusively textual) medium. It is generally considered as a biased, and reduced version of reality (Flewitt et al. 2009; Rowe 2012) that the researcher wants to study, or, more precisely, not of reality but of the data which have been collected. One may wonder whether this weakness primarily arise from the data collection itself. For example, on the one hand, linguistic and paralinguistic features of a spoken interaction can be precisely transcribed. But, on the other, an audio file captures only a limited part of the interaction: who is the exact addressee of a speech turn? What is the surrounding context? A video recording will bring much more information. However, in face-to-face situations, if a single camera is recording, important alternative perspectives may be missed. The "reality" of online multimodal interaction may be easier to capture than the face-to-face one. Actually a video-screen capture with audio-recording will accurately record the online context and participants' actions. It will not render the context and the action of an individual participant, but the shared context of all the participants. What is recorded into the video-screen capture is the common ground on which participants interact.

The need for shared transcription conventions

If individual studies are to build upon a common knowledge base, researchers need to be able to examine data across studies. For example, Sindoni (2013) studied participants' use of modalities in (non-educational) online environments that integrate audio, video, and text chat. She focused on what she termed "mode-switching," when a participant moves from speech to writing or the other way round. When analyzing transcriptions of video-screen online conversations, she observed that participants could be classified according to their preferred interaction mode (oral or written). She also observed that "[a]s anticipated, both speakers and writers, generally carry the interaction forward without mode-switching. This was observed in the whole video corpus" (Sindoni, 2013, §2.3.5). Hence she concluded that "those who talked did not write, and those who wrote did not talk. Turn-taking adheres to each mode."

In several CMILL situations that we studied, we had similar research questions. When analyzing data assembled in LETEC corpora such as Copéas or Archi21 (Mulce repository 2013), we observed that learners had a preferred mode of expression (oral or written), at least those at beginner level (Vetter and Chanier 2006). In contrast with Sindoni (2013), analyses of audiographic and 3D environments show that learners were mode-switchers (even modality- switchers). Choices of mode/modality depended on the nature of the task, and on the tutor's behavior (see Wigham and Chanier 2013b).

At this stage, one may expect that scientific discussions could take place between researchers studying online interactions, to debate contradictions, fine differentiations of situations, tasks, and so on. In order to allow this, data from the different approaches need to be accessible in standard formats, with publications clearly relating to data and data analyses, and explicit information given about the format of the transcriptions, their code, and transcription alignments with video. However Sindoni's data are not available. The inability to accumulate findings across CMILL studies or to contrast data from one study with other examples, available in open-access formats, is still holding back the scientific advancement of the CALL field. Overcoming these limitations will require researchers' development of common guidelines for transcription, coding, and analysis.

Transcription, coding, and analysis

Jenks (2011, 5) describes the main functions of transcription as the following: (1) represent; (2) assist; (3) disseminate and (4) verify. Transcription seeks to represent the interaction, that is, a multimodal discourse which it would have been impossible to analyze in its "live" state. Transcription assists the analysis which will be made of the data. Henceforth this research depends on the coding of interaction, whether this code will be compatible with analysis tools. Dissemination refers to the repeated process of analysis, either by the data compiler when s/he plans to publish several articles out of her/his corpus, or by the community. This function is essential to conform to the principle of rigor in scientific investigation (Smith, et al. 2011, 377):

> The fact of being able to store, retrieve, share, interrogate and represent in a variety of ways […] the results of one's analysis means that a semiotician can conduct a variety of analyses, and then explore the range of such analyses […]. Different analyses and perspectives upon analysis are encouraged, so that an analyst may produce multiple interpretations of a text.

The last function of transcription refers to the need of the academic community to know, as Jenks (2011, 5) put it, "whether any given claim or observation made is demonstrably relevant to the interactants and interaction represented in the transcript." Checking this claim requires procedures for estimating the level of agreement between transcribers.

The first function, representation requires deliberation because transcription represents only a partial view of interaction corresponding to phenomena the researcher wants to examine. Obviously, the richness of multimodal acts, as sketched in Figure 28.2, cannot be simultaneously transcribed to represent all of the multimodal perspectives captured through the video.[1] It is therefore important to adopt a methodological approach when transcribing and coding by taking into account the decision steps enumerated in Table 28.2.

The first step (first line of Table 28.2) is to choose appropriate software for integrating video- screen capture and transcription layers of verbal and non-verbal acts for every participant. The aim is to appreciate whether these products and events support L2 language development. Co-occurrence of these modes could be compared to a concert performance, and the transcription task to writing a music score, where all the instruments/modes can be read and interpreted after having been aligned (cf. time priority for General act features in Table 28.2).

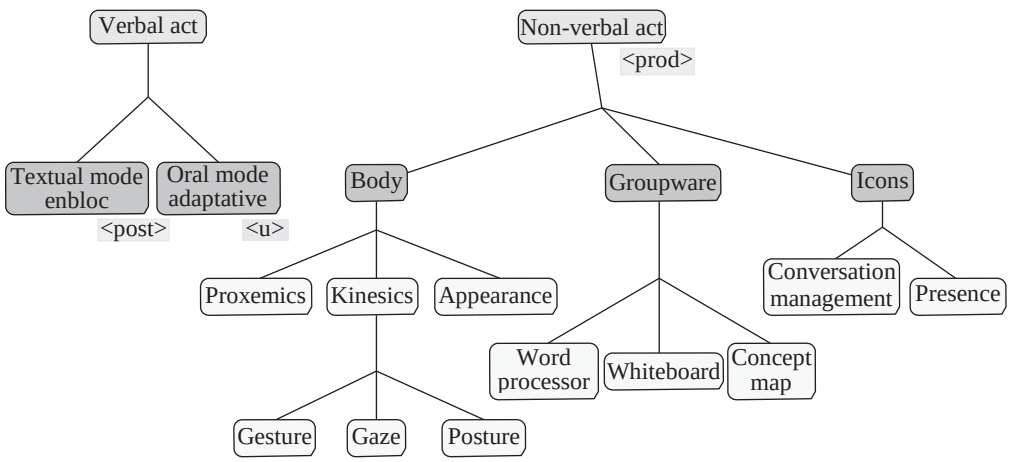

Figure 28.2 Multimodal acts as collected and studied within Mulce corpora repository (2013).

Table 28.2 Decision steps with their respective main features and comments for transcription and coding.

Decision steps	*Main features*	*Comments*
1. Choose software: integration of video screen capture and transcription layers	• Corresponding to varying participants and modes • Timeline: every layer aligned on the video	Avoid page-based transcriptions, Smith (ibid., 360) Numerous open access software
2. Adopt (extend) coding scheme	• Easily learnable • For speech: prefer standard (and extensible) existing codes • For non verbal: detail your code and make it publicly available	• Check code reliably applied by coders (Rowe, 2012, 204) • e.g., CHAT used in CHILDES Extend it for your specific needs. • e.g., (Saddour et al., 2011)
3. Prioritize when coding interactions: time description	• Beginning and ending times of an online session • Beginning and ending times of every interaction • Guarantee timeline continuity: code speech silences (not pauses) as act	Relationships with other course activities Allows: sequencing of acts, duration of each one, observing simultaneity / overlap of acts Priority to speech verbal acts with silences coding allow study of interplay between verbal and non-verbal modes
4. Select output format	• Standard format for transcription and coding (e.g. XML or TEI compliant)	Guarantee machine readability and automatic processing Use of interoperable software analysis Depends of the features of the transcription software

Once the layout of the music score is "printed," that is, once the different layers, corresponding to each participant and the various modes have been opened within the transcription software, with the video ready for alignment, the coding process can start. In synchronous multimodal CALL environment, the alignment is preferably made around the speech verbal acts.

The second step concerns speech transcription (cf. coding scheme in Table 28.2), that is by at least typing the words corresponding to the sound signal. Table 28.2, line 2, lists key points to take into consideration and related issues. A large part of them are shared by the community of speech corpora. Out of this transcription (including words in the textchat tools), many software packages are able to automatically calculate word tokens, work types, and mean length utterance. This computation can measure individual participant contribution: numbers of turns; floor occupied in sessions; vocabulary diversity, and so on.

A further step (cf. Output format in Table 28.2), specific to CMILL concerns, is to relate this methodology to a more abstract model, such as the interaction space one. It is as an extension of the Text Encoding Initiative (TEI), well known in corpus linguistics and other fields of Humanities. It encompasses text types such as manuscripts, theater, literature, poem, speech, film and video scripts. The TEI is specifically designed to accept different levels of annotations, each one corresponding to a specific type of analysis (e.g., morpho-syntactic, semantic, discursive). The IS model, under development, is designed for CMC in general, and includes multimodal discourse. It is the product of a European research work- group, TEI-CMC (Beißwenger, Chanier, and Chiari 2014). The CoMeRe repository (Chanier et al. 2014) provides access to corpora of various CMC genres (SMS, blogs, tweets, textchat, combinations of email-forum-textchat, and multimodal acts coming from 3D, audiographic environments).

Conclusion

The multimodal corpus-based approach to research on online language learning described in this chapter is not used widely in the study of CALL today. However, there is a need to expand the community of CALL researchers who adopt a corpus-based approach. Learning designs for CMILL need to take full account of the material and multimodal nature of the technical and learning resources involved in order to promote learning mediated through these resources. In order for research to improve CMILL, researchers need to investigate learners' performance in such interaction spaces. Sharing these data and documenting the processes through which they were created is a necessary step for deepening research on multimodal CALL. The LETEC corpus provides an example of how the field might move forward to investigate CMILL by collecting a corpus of CMILL data. It has served as a site for exploration of the various challenges facing the researcher with the transcription of these data, their coding in a standard way, and analysis. As researchers continue to work with this and other multimodal corpora, this area will continue to see advances that will help to improve CMILL.

Note

1 Collecting multimodal data and transcribing them are time-consuming processes. As regards transcriptions, estimated ratios vary from 15:1 for speech (i.e., 10 mn of audio requires 2.5 hours to transcribe it) to 23:1 for both speech and gesture (Rowe 2012).

References

Baude, O, Blanche-Benveniste, C,. et al. 2006. "Corpus Oraux, Guide des Bonnes Pratiques 2006." Accessed March 7, 2016. https://hal. archives-ouvertes.fr/hal-00357706

Beißwenger, Michael, Thierry Chanier and Isabella Chiari. 2014. "Special Interest Group on CMC of the Text Encoding Initiative (TEI) Consortium." Tei-c.org. Accessed March 7, 2016. http://wiki.tei-c.org/index.php/ SIG:Computer-Mediated_Communication

Budach, Gabriele. 2013. "From Language Choice to Mode Choice: How Artifacts Impact on Language Use and Meaning

Making in a Multilingual Classroom." *Language and Education*, 27: DOI:10.1080/09500782.2013.788188

Chen, Nian-Shing, and Yuping Wang. 2008. "Testing Principles of Language Learning in a Cyber Face-to-Face Environment." *Educational Technology & Society*, 11, no. 3: 97–113.

Chanier, Thierry, and Ciara Wigham. 2016. "Standardizing Multimodal Teaching and Learning Corpora." In *Language-Learner Computer Interactions: Theory, Methodology and CALL Applications*, edited by Hamel Marie-Jo, and Catherine Caws, 215–240. Amsterdam: John Benjamins. DOI:10.1075/lsse.2.10cha

Ciekanski, Maud, and Thierry Chanier. 2008. "Developing Online Multimodal Verbal Communication to Enhance the Writing Process in an Audio-graphic Conferencing Environment." *ReCall*, 20, no. 2. DOI:10.1017/ S0958344008000426

Codreanu, Tatiana, and Christelle Celik. 2013. "Effects of Webcams on Multimodal Interactive Learning." *ReCall*, 25, no. 1. DOI:10.1017/S0958344012000249

Coryell, Joellen E., and M. Carolyn Clark. 2009. "One Right Way, Intercultural Participation, and Language Learning Anxiety: A Qualitative Analysis of Adult Online Heritage and Non heritage Language Learners." *Foreign Language Annals*, 42, no. 3: 483–504.

Dahlberg, Giulia M., and Sangeeta Bagga-Gupta, S. 2013. "Communication in the Virtual Classroom in Higher Education: Languaging Beyond the Boundaries of Time and Space." *Learning, Culture and Social Interaction*, 2, no. 3: 127–142.

De los Arcos, Bea. 2009. Emotion in Online Distance Language Learning: Learners' Appraisal of Regret and Pride in Synchronous Audiographic Conferencing. PhD diss., The Open University. Accessed March 7. http:// oro.open.ac.uk/id/eprint/44076

De los Arcos, Bea, Jim A. Coleman, and Regine Hampel. 2009. "Learners' Anxiety in Audiographic Conferences: A Discursive Psychology Approach to Emotion Talk." *ReCALL*, 21, no. 1: DOI:10.1017/S0958344009000111

Develotte, Christine, Nicolas Guichon, and Caroline Vincent. 2010. "The Use of the Webcam for Teaching a Foreign Language in a Desktop Videoconferencing Environment." *ReCALL*, 22, no. 3: DOI:10.1017/ S0958344010000170

Dooly, Melinda, and Mirjam Hauck. 2012. "Researching Multimodal Communicative Competence in Video and Audio Telecollaborative Encounters." In *Researching Online Foreign Language Interaction and Exchange: Theories, Methods and Challenges*, edited by Melinda Dooly and Robert O'Dowd, 135–161. Bern: Peter Lang.

Dooly, Miranda, and Randall Sadler. 2013. "Filling in the Gaps: Linking Theory and Practice through Telecollaboration in Teacher Education." *ReCALL*, 25, no. 1. DOI: 10.1017/ S0958344012000237

Dyke, Gregory, Kristine Lund, Heysawn Jeong, Richard Medina, Daniel Suthers, Jan van Aalst, et al. 2011. "Technological Affordances for Productive Multivocality in Analysis." In *Proceedings of the 9th International Conference on Computer-Supported Collaborative Learning (CSCL)*, Hong-Kong, China, 454–461. Accessed March 7, 2016. https://halshs.archives- ouvertes.fr/halshs-00856537

Erben, Tony. 1999. "Constructing Learning in a Virtual Immersion Bath: LOTE Teacher Education through Audiographics." In *WORLDCALL: Global Perspectives on Computer- assisted Language Learning*, edited by Robert Debski and Mike Levy, 229–248. Lisse: Swets and Zeitlinger.

Flewitt, Rosie S., Regine Hampel, Mirjam Hauck, and Lesley Lancaster. 2009. "What Are Multimodal Data and Transcription?" In *The Routledge Handbook of Multimodal Analysis*, edited by Carey Jewitt, 40–53. London: Routledge.

Goodfellow, Robin, Ingrid Jefferys, Terry Miles, and Tim Shirra. 1996. "Face-to-Face Language Learning at a Distance? A Study of a Videoconference Try-Out." *ReCALL*, 8, no. 2: DOI:10.1017/S0958344000003530

Hampel, Regine. 2012. "Multimodal Computer- Mediated Communication and Distance Language Learning." In *The Encyclopedia of Applied Linguistics*. Hoboken, NJ: John Wiley & Sons, Inc. DOI:10.1002/9781405198431.wbeal0811

Hampel, Regine. 2014. "Making Meaning Online: Computer-mediated Communication for Language Learning" In *Proceedings of the CALS Conference 2012*, edited by Anita Peti-Stantić and Mateusz-Milan Stanojević, 89–106. Frankfurt: Peter Lang.

Ho, Caroline, M. L., Mark Evan Nelson, and Wolfgang K. Müller-Wittig. 2011. "Design and Implementation of a Student-generated Virtual Museum in a Language Curriculum to Enhance Collaborative Multimodal Meaning- making." *Computers and Education*, 57, no. 1: 1083–1097.

Hoff, Erika., ed. 2012. *Research Methods in Child Language: A Practical Guide*. Oxford: Wiley-Blackwell.

Hurd, Stella. 2007. "Anxiety and Non-anxiety in a Distance Language Learning Environment: The Distance Factor as a Modifying Influence." *System*, 35, no. 4: DOI:10.1016/j. system.2007.05.001

Jenks, Christopher J. 2011. *Transcribing Talk and Interaction*. Amsterdam: John Benjamins.

Jones, Rodney H. 2004. "The Problem of Context in Computer-mediated Communication." In *Discourse and Technology: Multimodal Discourse Analysis*, edited by Philip Levine and Ron Scollon, 20–33. Washington, DC: Georgetown University Press.

Jones, Rodney H., and Christoph A. Hafner. 2012. *Understanding Digital Literacies: A Practical Introduction*. New York: Routledge.

Kress, Gunther, and Theo Van Leeuwen. 2001. *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London: Arnold.

Kress, Gunther (2010) *Multimodality A Social Semiotic Approach to Contemporary Communication*. London: Routledge.

Lamy, Marie-Noëlle. 2012a. "Click if You Want to Speak: Reframing CA for Research into Multimodal Conversations in Online Learning." *International Journal of Virtual and Personal Learning Environments*, 3, no. 1: 1–18.

Lamy, Marie-Noëlle. 2012b. "Diversity in Modalities." In *Computer-Assisted Language Learning: Diversity in Research and Practice*, edited by Glenn Stockwell, 109–126. Cambridge: Cambridge University Press.

Lamy, Marie-Noëlle, and Rosie Flewitt. 2011. "Describing Online Conversations: Insights from a Multimodal Approach." In Décrire La *Communication en Ligne: Le Face-à-face Distanciel*, edited by Christine Develotte, Richard Kern, and Marie-Noëlle Lamy, 71–94. Lyon, France: ENS Editions.

Liègeois, Loïc. 2014. Usage des Variables Phonologiques dans un Corpus d'Interactions Naturelles Parents-enfant: Impact du Bain Linguistique et Dispositifs Cognitifs d'Apprentissage. PhD dissertation, Clermont University. Accessed March 7, 2016. https://tel.archives-ouvertes.fr/tel-01108764

Mackey, A., & Gass, S. M. 2005. *Second Language Research: Methodology and Design*. Abingdon: Routledge.

MacWhinney, Brian. 2000. "The CHILDES Project: Tools for Analyzing Talk: Volume I: Transcription Format and Programs, Volume II: The Database." *Computational Linguistics*, 26, no. 4: 657–657.

McAndrew, Patrick, Sandra P. Foubister, and Terry Mayes. 1996. "Videoconferencing in a Language Learning Application." *Interacting with Computers*, 8, no. 2: 207–217.

Monteiro, Kátia. 2014. "An Experimental Study of Corrective Feedback during Video- Conferencing." *Language Learning and Technology*, 18, no. 3: 56–79.

Mulce Repository. 2013. "*Repository of Learning and Teaching (LETEC) Corpora*. Clermont Université: MULCE.org." Accessed March 7, 2016. http://repository.Mulce.org

O'Halloran, Kay L., Alexey Podlasov, Alvin Chua, and K. L. E. Marissa. 2012. "Interactive Software for Multimodal Analysis." *Visual Communication*, 11, no. 3: 352–370.

O'Halloran, Kay L., and Bradley A. Smith, eds. 2011. *Multimodal Studies: Exploring Issues and Domains*. New York: Routledge.

Reffay, Christophe, Marie-Laure Betbeder, and Thierry Chanier. 2012. "Multimodal Learning and Teaching Corpora Exchange: Lessons learned in 5 years by the Mulce project." *International Journal of Technology Enhanced Learning (IJTEL)*, 4, no. 1–2. DOI:10.1504/ IJTEL.2012.048310

Reinhardt, Jonathon, and Julie Sykes. 2014. "Special Issue Commentary: Digital Game and Play Activity in L2 Teaching and Learning." *Language Learning & Technology*, 18, no. 2: 2–8.

Royce, T. D. 2007. "Multimodal Communicative Competence in Second Language Contexts." In *New Directions in the Analysis of Multimodal Discourse*, edited by Teddy D. Royce and Wendy Bowcher, 361–390. Mahwah, NJ: Lawrance Erlbaum.

Rowe, Meredith. L. 2012. Recording, Transcribing, and Coding Interaction. In *Research Methods in Child Language: A Practical Guide*, edited by Erika Hoff: Oxford: Wiley- Blackwell. DOI:10.1002/9781444344035.ch13

Satar, Müge H. 2013. "Multimodal Language Learner Interactions via Desktop Conferencing within a Framework of Social Presence: Gaze." *ReCALL*, 25, no. 1: DOI:10.1017/ S0958344012000286

Schwienhorst, Klaus. 2002. "The State of VR: A meta-analysis of Virtual Reality Tools in Second Language Acquisition." *Computer Assisted Language Learning*, 15, no. 3: DOI:10.1076/call.15.3.221.8186

Segouat, Jérémie, Ammick Choisier, and Amelies, Braffort. 2010. Corpus de Langue des Signes: Premières Réflexions sur leur Conception et Représentativité. In *L'exemple et Le Corpus Quel Statut? Travaux Linguistiques du CerLiCO*, 23: 77–94. Rennes: Presses Universitaires de Rennes.

Shih, Ya-Chun. 2014. "Communication Strategies in a Multimodal Virtual Communication context." *System*, 42: DOI:10.1016/j. system.2013.10.016

Sindoni, Maria Grazia. 2013. *Spoken and Written Discourse in Online Interactions: A Multimodal Approach*. New York: Routledge.

Smith, Bradley A., Sabine Tan, Alexey Podlasov, and Kay L. O'Halloran. 2011. "Analyzing Multimodality in an Interactive Digital Environment: Software as Metasemiotic Tool." *Social Semiotics*, 21, no. 3: 359–380.

Stickler, Ursula, Caroline Batstone, Annette Duensing, and Barbara Heins. 2007. "Distant Classmates: Speech and Silence in Online and Telephone Language Tutorials." *European Journal of Open, Distance and e-Learning*, 10, no. 2. Accessed January 5, 2017. http://www. eurodl.org/

Stivers, Tanya, and Jack Sidnell. 2005. "Multimodal Interaction." Special Issue of *Semiotica*, 156, no. 1–4: 1–20.

Tayebinik, Maryam, and Marlia Puteh. 2012. "The Significance of Self-esteem in Computer Assisted Language Learning (CALL) Environments." *Procedia – Social and Behavioral Sciences*, 66: DOI:10.1016/j. sbspro.2012.11.294

Vetter, Anna, and Thierry Chanier. 2006. "Supporting Oral Production for Professional Purpose, in Synchronous Communication With heterogeneous Learners." *ReCALL*, 18, no. 1. DOI:10.1017/S0958344006000218

Wang, Yuping. 2004. "Supporting Synchronous Distance Language Learning with Desktop Videoconferencing." *Language Learning & Technology*, 8, no. 3: 90–121.

Wang, Yuping. 2006. "Negotiation of Meaning in Desktop Videoconferencing-Supported

Distance Language Learning." *ReCALL*, 18, no. 1: DOI:10.1017/S0958344006000814

Wang, Yuping. 2007. "Task Design in Videoconferencing-Supported Distance Language Learning." *CALICO Journal*, 24, no. 3: 591–630.

Wang,Y., Chen, N-S., and Levy, M. 2010 "Teacher Training in a Synchronous Cyber face-to-face Classroom: Characterizing and Supporting the Online Teachers' Learning Process. " *Computer-Assisted Language Learning*, 23, no. 4: 277–293.

Wang, Yuping, and Nian-Shing Chen. 2012. "The Collaborative Language Learning Attributes of Cyber Face-to-face Interaction: The Perspectives of the Learner." *Interactive Learning Environments*, 20, no. 4: 311–330.

Wigham, Ciara, and Thierry Chanier. 2013a. "A Study of Verbal and Non-verbal Communication in Second Life – the ARCHI121 experience." *ReCall*, 25, no. 1. DOI: 10.1017/S0958344012000250

Wigham, Ciara, and Thierry Chanier. 2013b. "Interactions Between Text Chat and Audio Modalities for L2 Communication and Feedback in the Synthetic World Second Life." *Computer Assisted Language Learning*, 28, no. 3. DOI:10.1080/09588221.2013.851702

Yang, Shu Ching. 2001. "Language Learning on the World Wide Web: An Investigation of EFL Learners' Attitudes and Perceptions." *Journal of Educational Computing Research*, 24, no. 2: 155–181.