

## SHARING A MULTIMODAL CORPUS TO STUDY WEBCAM-MEDIATED LANGUAGE TEACHING

Nicolas Guichon, [Université Lumière Lyon 2](#), [Laboratoire ICAR](#)

This article proposes a methodology to create a multimodal corpus that can be shared with a group of researchers in order to analyze synchronous online pedagogical interactions. Epistemological aspects involved in studying online interactions from a multimodal and semiotic perspective are addressed. Then, issues and challenges raised by corpus creation and sharing are examined with a particular focus on ethics. Basing my discussion and analysis around a particular research project, the steps involved in the creation of a multimodal and shareable corpus are described and the scientific benefits of a collective exploration of data and their subsequent multimodal representations are discussed and illustrated. It is finally claimed that corpus sharing can contribute vastly to the field of computer-assisted language learning by enhancing its scientific robustness as it favors a multidisciplinary, systematic, and in-depth analysis of multimodal data.

**Language(s) Learned in this Study:** French

**Keywords:** Corpus, Distance Language Learning and Teaching, Telecollaboration, Research Methods

**APA Citation:** Guichon, N. (2017). Sharing a multimodal corpus to study webcam-mediated language teaching. *Language Learning & Technology*, 21(1), 55–74. Retrieved from <http://llt.msu.edu/issues/february2017/guichon.pdf>

**Received:** November 8, 2015; **Accepted:** July 19, 2016; **Published:** February 1, 2017

**Copyright:** © Nicolas Guichon

### INTRODUCTION

Webconferencing-based second language teaching continues to gain momentum as more courses are delivered using such tools (Kern, 2014). In order to describe and understand what is at stake in a pedagogical interaction that brings together language teachers and language learners via a webconferencing platform, computer-assisted language learning (CALL) researchers may benefit from sharing different theoretical and methodological perspectives on the same dataset.

The present paper is one of the outcomes of ISMAEL, a collective project which brings together twenty specialists in technology mediated language teaching, interactions, and semiotics. The ISMAEL corpus (Guichon, Blin, Wigham, & Thouësny, 2014) is composed of data from a six-week telecollaboration project between teacher trainees and learners of French. It is made up of audio and video recordings of all the exchanges between teachers and trainees, group debriefing sessions with trainee teachers, and interviews with learners.

To qualify as a corpus, these data were structured, transcribed, annotated, and contextualized. While most of these operations involving corpus creation are now widespread in linguistics (see Knight, Evans, Carter, & Adolphs, 2009), the interest of the present paper lies in the scientific stance towards building a multimodal corpus to inform given phenomena pertaining to the field of CALL and to sharing that corpus with a community of researchers. I contend that the aggregation of data into a coherent whole and the issues that are raised by its curation, sharing, and analysis are fundamental activities at the core of digital humanities (Burdick, Drucker, Lunenfeld, Presner, & Schnapp, 2012) that need to be explored further by the CALL field.

This paper proposes a reflection organized into two parts. First, the epistemological aspects involved in

studying online interactions from a multimodal and semiotic perspective will be addressed along with issues raised by corpus creation and sharing. Then, taking the ISMAEL project as a guiding thread, I will explore the steps involved in the creation of a multimodal and shareable corpus and show how the organization of a data session can make a powerful contribution to the socializing of a corpus. With such a description, the article aims to better understand the interaction between science and technique since any research process produces *intermediary objects* (Vinck, 1999) that are of interest when one wishes to examine the scientific practices related to the creation of corpora and their sharing.

## THEORETICAL FRAMEWORK

Although different theoretical frameworks and methodological approaches are already being used by researchers to explore the ISMAEL corpus, the framework chosen for this paper is rooted in the field of multimodal analysis of interactions (Jewitt, 2009; Norris, 2004; Sindoni, 2013). This semiotic approach will be briefly presented before issues involved in multimodal corpus creation and sharing are discussed.

### A Semiotic Approach to Webcam-mediated Language Teaching

Most CALL research to date on webcam-mediated exchanges has focused on learners and has endeavored to study how they manage tasks in different conditions, usually comparing audio- and webconferencing (e.g., Guichon & Cohen, 2014; Yamada & Akahori, 2009; Yanguas, 2010). Such studies are usually experimental, punctual (participants are put in a situation that is unfamiliar), and seldom describe the activity itself but rather how participants report their online activity through post-task questionnaires or interviews.

In parallel, an increasing number of researchers have been turning their interest towards online teacher activity and have started to take semiotic dimensions into account. This small but growing body of research usually draws on case studies and a limited collection of data. For instance, Satar (2012) includes screen recordings of webcam-mediated interactions between language learners and identifies five types of gaze involved in webconferencing. Develotte, Guichon, and Vincent (2010) concentrate on the use of the webcam during pedagogical interactions of five teacher trainees. Their study recommends teacher training which focuses on developing online teachers' capacity to utilize their own image depending on pedagogical objectives and learner needs.

Because online teacher activity is mediated by an array of technologies (in particular a webcam, a microphone, and text chat), one possible approach to studying this pedagogical situation is to focus on the multimodal aspects and uncover the many facets of the online teachers' *semio-pedagogical activity* (Guichon, 2013). The semio-pedagogical dimension of teachers' activity refers to their use of the various semiotic and technological resources and to the degree of professional competence they display in order to facilitate language learning.

I contend that multimodal research (Jewitt, 2009) offers an ideal framework to study online teacher activity for three reasons. First, it allows one to take into account the *whole* repertoire of semiotic resources individuals use when they are engaged in an activity, and not only verbal language. Then, it includes in its analyses the process through which semiotic resources progressively contribute to meaning-making to enrich a socially shared repertoire. Lastly, it looks closely at the ways in which the semiotic modes are combined or dissociated in what Norris (2004) has called multimodal orchestration. At a micro level, the analysis of semio-pedagogical competence could thus include a moment-to-moment inspection of online teachers' activity when they are performing pedagogical actions (e.g., giving instructions, providing feedback, etc.) while using the semiotic resources at their disposal (their voice, facial expressions, gestures) in the specific context of a webconferencing-based interaction.

I suggest that a corpus-based approach makes the study of the development of semio-pedagogical competence possible as it enables the analysis of a series of interactions involving the same participants.

Researchers may then examine whether the teachers' management of the online synchronous interaction becomes progressively more appropriate and whether their interactional repertoire becomes richer. The tenets for building such corpora are presented in the next section.

### Principles Guiding the Creation of the ISMAEL Corpus

Over the last decade, more and more researchers have called upon “a new generation of corpora” making it possible for linguists “to gain a more comprehensive view of the characteristics of language ‘beyond the text’” (Knight et al., 2009, p. 2). Besides, new communication and pedagogical practices (audioconferencing and videoconferencing) have created the need for CALL researchers to build corpora, gathering data in the different modes used by the actors involved in such synchronous mediated interactions. Indeed, a corpus-based approach provides the means of following the same individuals interacting over time and enables researchers to “reveal the inventive ways they are able to co-construct meaning with their interactional partners” (Seedhouse, 2005, p.265).

Basing the present analysis around the ISMAEL project, I describe what is at stake when trainee teachers participate in online pedagogical interactions with language students. The research aim is to gain insights into the specificities of this techno-pedagogical situation and, ultimately, inform teacher education (Hampel & Stickler, 2012). Some of the questions that were explored by the researchers involved in the project can be formulated as follows: How do second language teachers in training harness semiotic and pedagogical resources in order to teach a foreign language online, more precisely to (a) engage students in meaningful and medium-appropriate exchanges, (b) manage feedback during the interactions, (c) manage communication breakdowns, (d) provide help to their students, (e) negotiate meaning, (f) provide rhythm to the exchange, and (g) enhance their online teaching presence? (see Guichon & Tellier, 2017).

As the preceding list of research questions demonstrates, the focus of the research is placed on what actually happens rather than what should happen. In other words, it is descriptive rather than prescriptive. In order to investigate these questions, a corpus-based approach was adopted to provide researchers with a collection of data that were representative of the pedagogical situation under study. Four principles presided over the reflection on data collection in order to augment the validity of the approach:

1. *Multimodal quality*: Because of the semiotic focus of the project, it was important to collect multimodal data that would provide information on the ways in which teachers and learners interacted in the verbal, visual, and written modes. Special attention was paid to ensure that sound and image quality of the captured online interactions was good enough for future multimodal enquiry (for an example, see [Clip 1](#)).
2. *Contextual integrity*: It was decided that the ecology of the pedagogical situation around which the research protocol was designed should be altered as little as possible. All the captured data were included even if technical difficulties or other inevitable mishaps (e.g., absences) occurred. While these types problems made the data harder to interpret, a wealth of details that experimental approaches are bound to overlook were kept in the data.
3. *Developmental quality*: Since the ISMAEL project aimed to study how novice teachers become attuned to teaching online, it was decided that the data would be collected over several consecutive weeks. Because it spanned six weeks, the ISMAEL data cannot be considered to be longitudinal per se but nonetheless provide enough opportunities to capture actors' trajectories encompassing the discovery of new tools to the development of certain teaching practices.
4. *Participants' diversity*: Finally, it was important to include as many novice teachers as possible in the study because some features (e.g., age, gender, familiarity with computers, professional experience) could have an impact on the development of semio-pedagogical competence.

In sum, the ISMAEL project involved putting together a corpus that could capture pedagogical

interactions in their multimodal diversity and integrity and which would include a large enough collection of data involving several actors and spanning a significant period of time.

### Data Sharing and Shareability

The approach of this project to corpus creation was in line with that of Chanier & Wigham (2016) who proposed the Learning and Teaching Corpora (LETEC) methodology, defined as a systematic collection and structuration of all the data from interactions that occur online and have an educational purpose. These authors explain how interactions can be (1) structured and standardized thanks to an XML protocol, (2) described with metadata and (3) made accessible on a repository in order to facilitate sharing with other researchers. Following Chanier & Wigham's reflection, this section concentrates on the issue of data sharing which deserves a reflection closely linking practical and scientific aspects.

Data sharing becomes possible—or desirable—only if a certain number of conditions have first been fulfilled. I propose to define *shareability* as the result of the practical, ethical, epistemological and social conditions that need to be met in order to enhance the possibility given to a group of researchers to share a dataset. Chanier and Ciekanski (2010) have underlined that the costs of producing shareable corpora can become acceptable only if the perspective of sharing becomes real.

The trend towards data sharing (for their pioneering collective work on learner corpora, see Granger, Gilquin, & Meunier, 2015) implicates new scientific responsibilities: researchers and technical support involved in such ventures need to ensure the quality of the corpus itself (image, sound), its interoperability (making it possible to access the data with different operating systems and software), and its documentation (providing future collaborators with information about when and how the data were collected, who the participants were, and how ethical issues were addressed). In sum, because it brings real value to a dataset, attention devoted to corpus quality, access, and documentation are important steps to facilitate the re-use of data by other researchers for further analyses or even replication of published studies.

For the CALL research community, sharing data presents a host of practical and scientific benefits. From a practical perspective, sharing allows the reduction of human and financial costs involved in the different stages of corpus creation. It especially permits the division of the time-consuming process of transcription and annotation (see section below on Core Data) among the researchers working on the same corpus. This allows more data to be processed, thus widening the empirical basis and scope of subsequent studies. From an epistemological perspective, sharing allows the corpus to be approached from different research perspectives, theoretical standpoints, and methodological approaches (Chanier & Ciekanski, 2010). More often than not, articles that are published in the field of CALL focus on data that have been gathered from a project in which the author was involved, which can create various biases. As underlined by Cappeau and Gadet (2007), the relationship of one researcher with a given context and his or her involvement in the research design and data collection may have a strong influence on the interpretation of the data, which can eventually be detrimental to the proposed analyses. I argue that there can be real benefits from gathering insiders (researchers who were closely involved in the project) because they can bring important insights about contextual and individual factors as well as outsiders (researchers who do not know the participants and the context) who can look at the data from a fresh perspective.

Yet, despite its obvious interest for science, data sharing and re-use remain a rare practice in language research (Dybkjær & Ole Bernsen, 2004) and CALL research (Dooly, 2015). It may represent a cost, working with data that were gathered by others in unfamiliar contexts, that few researchers are willing to bear. However, the main obstacle to data sharing has been pinpointed by Lemke (2013) who deplores the permanence of “a strong individualist bias in our modernist traditions of research” which leads us “to define our objects of study in such a way that a single researcher could in principle come to understand them” (p. 287).

I contend that making data sharing easier and more desirable requires the pursuit of different methodological strategies. The focus on shareability has been one defining aspect of the ISMAEL project. The rest of this article explores the different steps involved in the process of corpus creation and sharing.

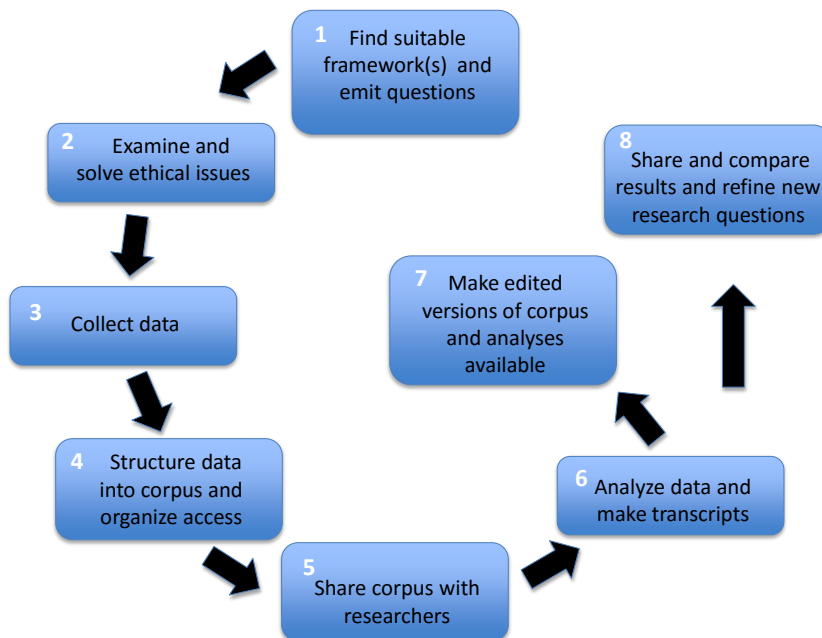
## METHODOLOGICAL ASPECTS INVOLVED IN THE CREATION OF A MULTIMODAL AND SHAREABLE CORPUS

### Global Presentation of the ISMAEL Project and its Context

In order to provide readers with elements pertaining to the context of the pedagogical situation that served as the source of data, a quick description of the training program and the participants is now proposed. The context for this study was a telecollaboration project that brought together French as a foreign language (FFL) trainee teachers and undergraduate business students in their third semester of learning French. Each trainee teacher was randomly assigned to one or two business students (see [Figure 3](#) for an overview). The trainees were second-year students of the Master of Arts program in Teaching French as a Foreign Language at Université de Lyon 2. The telecollaboration project formed part of an optional module entitled *Learning to Teach Online*. The main objectives consisted of developing hands-on professional skills to teach FFL online by planning and mediating online interactions. For the lower intermediate (B1) language students from Dublin City University (DCU), the telecollaboration project formed part of a 12-week blended module entitled *French for Business*.

The participants met weekly for six 40-minute online sessions in autumn of 2013 via the webconferencing platform Visu (Guichon, Bétrancourt, & Prié, 2012). Each online session focused on Business French objectives and ultimately prepared the students to apply for an internship in France. From the start, this telecollaboration project was also envisaged by the two instructors on both sides as an opportunity to research online language learning and teaching. Such an intimate link between training and research purposes created the basis for a fruitful collaboration, the involvement of researchers at both ends ensuring their equal commitment to corpus creation.

[Figure 1](#) presents the eight stages of corpus-based research that favors a semiotic approach and is strongly dependent on data sharing. All the steps will be dealt with in the rest of this paper, with varying degrees of depth, and the ISMAEL project will serve to illustrate how this methodological approach has been implemented. From the emission of research questions to the dissemination of analyses, the article will endeavor to show (1) why ethical issues have a crucial impact on many aspects of the process, (2) what steps are involved in the creation of a shareable corpus, and (3) how a community can be built around a corpus to explore online teaching from a semiotic perspective.



*Figure 1.* Stages of a collective and qualitative corpus-based research.

### Dealing with Ethical Issues

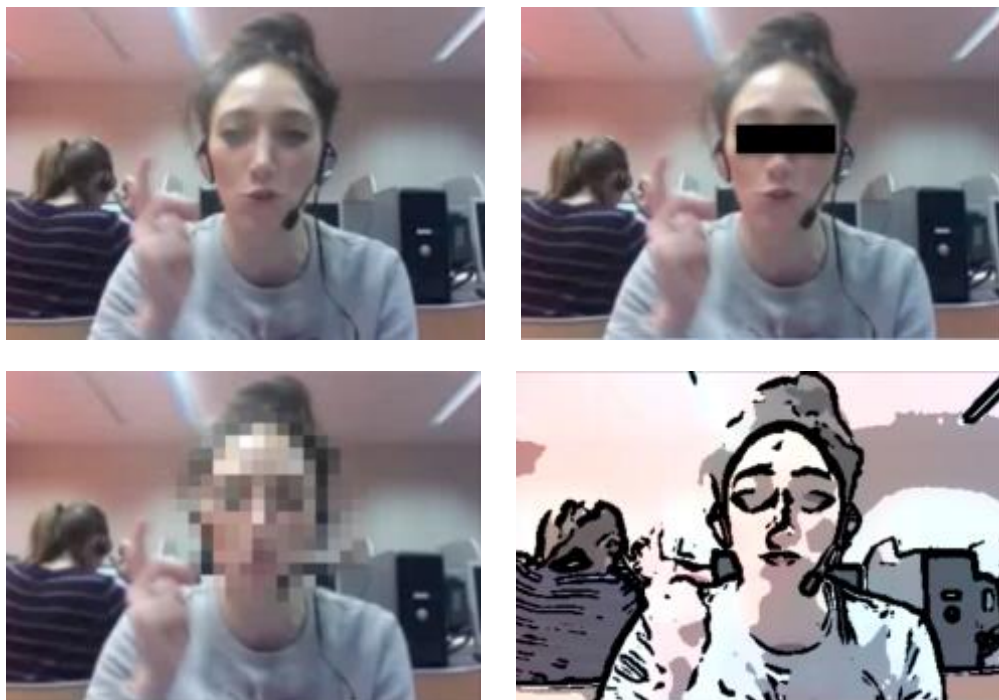
More and more papers are written about the ethical dimensions involved in Internet-mediated research (IMR) in order to help researchers “with the process of ethical decision making in the context of specifying and implementing appropriate IMR research designs” (British Psychological Society, 2013, p. 2). In this section I review the ethical considerations involved in collecting, sharing, and displaying multimodal data. These three aspects of a research project are highly sensitive since data are collected in an institutional setting and their prime interest for subsequent multimodal analyses comes from their visual nature, making the issues of confidentiality and privacy especially problematic.

As stated by the British Psychological Society’s Ethics guidelines (2013), “A researcher should be clear about the extent to which their own collecting and reporting of data obtained from the internet might pose additional threats to privacy over and above those that already exist, and whether this might expose participants to potential harm of any sort” (p. 10). For instance, in the case of the ISMAEL project, the data involve teachers who were learning to teach online and could sometimes be clumsy; thus online clips displaying them in awkward situations can potentially harm their future employability. Furthermore, the language students involved in the telecollaboration project needed to be assured that withdrawing their consent would not have any impact on their course assessment. This was actually ensured by the fact that the researcher in Dublin did not know what the participants had decided about their participation until she had completed her students’ evaluation. Ethical considerations have an impact on several aspects of the research process as summarized in [Table 1](#). I go through the different steps encountered in the ISMAEL project to solve ethical issues.

Table 1. *Ethical Issues*

<b>Aspect</b>	<b>Possible steps</b>
<b>Data collection</b>	<ul style="list-style-type: none"> <li>• Go through the ethics procedures in the institutions where the participants study (or work) before the data collection is set to begin</li> <li>• Design consent forms accordingly stating the research aims and how data will be used in published research, oral presentations and training material and inform participants of the levels of risk to the confidentiality of their data</li> <li>• Explain the aims of the research to participants and, if possible, examples of publications similar to those expected from the current research project, and offer opportunities to ask questions</li> <li>• Get all participants to sign the consent forms and clarify that they can act on their rights to withdraw consent at any time (see <a href="#">Appendix A</a>)</li> </ul>
<b>Data structuration</b>	<ul style="list-style-type: none"> <li>• Put only one or two researchers in charge of the raw data and protect them with passwords</li> <li>• Keep in the data only those of the individuals who have agreed to participate in the research project</li> </ul>
<b>Corpus sharing</b>	<ul style="list-style-type: none"> <li>• Write a protocol stating what researchers involved in the project can and cannot do with the data, get them to sign the protocol and store the signed agreements as part of the data</li> <li>• Specify standardized corpus referencing in the protocol</li> </ul>
<b>Transcripts and Dissemination</b>	<ul style="list-style-type: none"> <li>• Ensure the transcripts and research outputs as they are disseminated are compliant with the consent forms as signed by participants and ethics procedures</li> </ul>

There are different ways of anonymizing the participants' image in the data, their eyes usually being the most revealing feature of their identity. If we take the example of one trainee teacher (who did give her consent for this research), her eyes can either be covered with a black patch or her face can be pixelated. Another possibility is to use software capable of transforming the image into a drawing.



*Figure 2.* Different ways of anonymizing participants' data.

As exemplified in [Figure 2](#), the different ways of anonymizing the participants' distinctive features do not provide the same apprehension of the data. For example, as Knight et al. (2009) remarked pixelating faces or using drawings “can blur distinctions between gestures and language forms rendering the data unusable for certain lines of linguistic enquiry” (p. 8). Since a semiotic approach was adopted for this research, facial expressions were of crucial importance. Thus only those participants (7 teachers and 12 learners) who had consented to having their images examined by researchers were included in the shared corpus.

To conclude on the subject of ethics, it seems important to insist that ethics concerns not only how participants are treated in a research project but also how *an ethics of sharing* is encouraged among a group of researchers. Thus, for the ISMAEL project, a document was drawn up to state the researchers' rights and obligations with the data. It included

- information relating to the access of the data,
- expectations concerning the respect of the participants' ethical wishes (anonymization of data),
- expectations concerning the obligation to add new annotations to the corpus in order to pool them with the rest of the group, and
- expectations concerning the explicit and complete citation of the corpus.

Ensuring that all researchers sign such a document is a useful way to formalize how everyone commits to the group and to spell out the meaning of data sharing. In brief, data sharing entails mutual responsibility: respect from the researchers to the participants and also respect among researchers.

## THE CREATION OF A MULTIMODAL CORPUS

A prime reason for creating a corpus which brings together all the data from a project is to ensure long-term accessibility for researchers. The Visu platform which was used for the telecollaboration project in 2013 and stored all the online exchanges no longer exists. Retrieving all the video files and structuring them into a corpus became the safest means to ensure that the project could be sustained beyond the life of the teaching platform. In the following section, I explain how two categories of data were gathered to



form a complex corpus and how special attention was paid to making these data as shareable as possible.

### Core Data

In order to have access to the trainee teachers' online activity, the core research data were made up of the traces of the interactions that were automatically collected and stored by Visu. To provide an illustration of the data, [this 8-second clip](#) shows three participants—one teacher trainee (Adèle, on the left of the screen) and two learners from DCU (Alannah at the top and Catriona)—involved in a role playing activity where the two learners had to plan a birthday party and suggest their ideas to their manager (role played by the trainee teacher). Adèle is trying to encourage the two learners to speak together (*Try to talk, both of you, it's your project*) while she makes it clear she wants to remain in the background and play the role of the manager (*and I'm grading you, I am the manager, I'm judging you*).

Adopting a semiotic approach entails paying attention to all the multimodal elements that make up the interactions. Software such as [ELAN](#) (Sloetjes & Wittenburg, 2008) can help researchers annotate the data. The [second clip](#) shows the same extract as Clip 1 but inserted into ELAN and annotated thanks to the functionalities offered by this software. ELAN's functionalities include the possibility to replay a video extract repeatedly, a timeline aligned with media time that facilitates the navigation of the data, and the possibility of creating tiers to annotate each new element under scrutiny. This usually starts with the transcription of the verbal utterances of the participants and then any other elements (smiles, facial expressions, proximity to the screen) that are deemed relevant to the study. Although time-consuming to learn to use efficiently, such software is precious for exploring minutely how verbal and co-verbal resources interact, and on large data sets, to identify patterns in performance (for instance the production of smiles at certain points of the interactions, or a change in the ways teachers provide negative feedback over the period of the telecollaboration).

### Complementary Data

The minute examination of verbal and co-verbal resources could provide invaluable insights as to how teachers manage an online interaction. Nevertheless, because CALL research seeks to bring insights that are valuable to the field of education, I contend that such micro analyses should be complemented with actors' perceptions, institutional constraints, and technological constraints because they also contribute to shaping pedagogical interactions. The approach of the ISMAEL project thus not only involves an analysis of the recorded online interactions but combines it with the study of complementary data which make up "a dynamic constellation of resources" (Flewitt, Hampel, Hauck, & Lancaster, 2009, p. 44) in an effort to enhance the social validity of research that dwells on description but also seeks to devise possible modes of intervention. We follow Jones (2004) when he cautions researchers working in the computer-mediated communication (CMC) field not "to stop at the screen's edge" (p. 24) so as to gain a fuller understanding of context and not to separate what is going on online from what is going on around the participants. This is why the lesson plans that are included in the corpus can prove crucial for outsiders to understand the situation.

Complementary data in the ISMAEL project are comprised of the following:

- lesson plans to understand the pedagogical objectives pursued by the teachers;
- transcriptions of group debriefing sessions with teacher trainees giving insight as to their pedagogical intentions and perceptions the day after each interaction;
- field notes and photographs gathered during the interactions; and
- post-interviews with teacher trainers, teacher trainees, and language students to examine some aspects under scrutiny (e.g., how trainees reported how they provided feedback, how language students considered the interactions with their online teachers).

## From an Array of Multiple Data to a Shareable Corpus

Multiple data can only qualify as a corpus once they have undergone a certain number of operations (i.e., annotation, structuration, contextualization, and accessibility) that are detailed in this section. One of the most time-consuming and costly operations involves *annotating* the data—that is, transcribing the verbal utterances of the participants and annotating the exchanges according to a coding scheme (see [Clip 2](#) and the section on Core Data above). Adolphs and Varter (2013) point out the interest of developing coding schemes for transcription so that “they can be shared across different research communities and with different community cultures and different representational and analytical needs” (p. 155). Since one of the aims of the ISMAEL project was to analyze several novice teachers’ trajectories, it was thought essential to annotate the same session for all teachers and follow several teachers over six weeks (see [Figure 3](#)). Transcribed online data for the ISMAEL project totaled 15 hrs 23 mins and involved the work and time of Benjamin Holt, a doctoral student who was in charge of devising a coding scheme for transcribing the verbal exchanges, training up to ten research assistants to use ELAN, supervising their work, and ensuring the quality and the consistency of their annotations. Then, as Chanier and Ciekanski (2010) insist, data have to be *structured*, that is, each piece of the corpus has to be digitalized, labeled, and organized so that the researchers involved in the project can easily retrieve the elements that are needed for their analyses. Next, the data have to be *contextualized* with the provision of all the information required to understand fully who the participants were and in what pedagogical and technical environments the interactions took place as exemplified in [Table 2](#).

**Table 2.** Complementary information to contextualize primary data

Types of complementary information	Information
Socio-demographics and profiles	<ul style="list-style-type: none"> <li>• age and gender of the participants</li> <li>• familiarity with tools used for the interaction</li> <li>• level in target language and motivations</li> <li>• experience in learning or teaching online</li> </ul>
Pedagogical dimensions	<ul style="list-style-type: none"> <li>• nature of tasks and themes</li> <li>• documents used</li> <li>• instructions given</li> <li>• place of the interactions within the curricula</li> </ul>
Temporal dimensions	<ul style="list-style-type: none"> <li>• length of each interaction, frequency of interactions (e.g. once a week)</li> <li>• duration of module (e.g. a semester)</li> </ul>
Technological dimensions	<ul style="list-style-type: none"> <li>• type of software and hardware used (e.g. desktop or laptop, devices used for recording, etc.)</li> </ul>

Finally, an additional semiotic layer can be useful to increase the *accessibility* of the data once they have been uploaded to an online repository. Even though researchers are more and more comfortable with databases, finding one’s way into a complex corpus can be daunting if one has to go through a list of long and confusing labels. For the ISMAEL project, it was decided to implement an html portal protected by a password to organize the access to the corpus repository as seen in [Figure 3](#). The interface works as a table with two organizing entries: the online teacher trainees with their students (in ordinate) and the six sessions with a summary of their content (in abscissa).

Teacher trainees	Learners		Session 1 October 15 <sup>th</sup> French professional world	Session 2 October 22 <sup>nd</sup> Professional experience	Session 3 November 12 <sup>th</sup> Internship in Reims	Session 4 November 19 <sup>th</sup> Project management	Session 5 November 26 <sup>th</sup> Project implementation	Session 6 December 3 <sup>rd</sup> Job interview
Adèle*	Alannah	Catrina*	✓ 43m48s	✓ 40m55s	✓ 33m03s	✓ 34m27s	✓ 41m04s	✓ 35m55s
Emilie*	Aiden	Fiona	✓ 22m50s	✓ 33m00s		✓ 25m37s		✓ 24m12s
Samia*	Sean*	Angela*	✓ 30m20s	Sean absent ✓ 34m34s	⊕ ✓ 31m50s	✓ 35m20s	✓ 39m46s	X
Etienne*	Conor	Sophie				✓ 20m57s		
Méllissa*	Ana*	Alejandra	✓ 42m11s	✓ 39m24s	✓ 37m03s	✓ 30m13s	✓ 32m28s	✓ 35m53s
Séverine*	Naomi			X		✓ 34m47s		
Victor*	Liam		✓ 33m25s	Session with Conor ✓ 27m46s	✓ 29m18s	✓ 34m00s	⊕ ✓ 29m01s	X

✓ Session transcribed  
 X Session unavailable  
 \* Interview  
 ⊕ Exterior view  
 Lesson plan  
 Debriefing the next day  
 Total time transcribed: 15h, 23m, 07s

Figure 3. The interface to access the ISMAEL corpus.

An interface like the one displayed in Figure 3 functions as a gateway to the structured data behind it. The whole corpus can thus be seen at a glance displaying what data are available for each session and for each group (video clips and verbal transcription) and the length of each interaction. The icons (detailed at the bottom of Figure 3) are hyperlinks that give access to either core or complementary data. Pictures (either the original images or conversions of these to line drawings) also remind researchers what the participants have agreed to in their consent forms so that data can be used accordingly, knowing from the outset whether or not they will be able to show participants’ faces clearly.

Collecting data, structuring them into a corpus, providing contextual information, and facilitating their accessibility for research are thus the different steps involved in the creation of a shareable corpus. All these steps necessitate time and effort, efficient and strategic decision-making procedures, and a constant attention to ethical issues and to the researchers’ needs in order to maximize the interest and shareability of the corpus. Yet, creating such a corpus is not an end in itself and the potential for sharing has to lead to real sharing and scientific exploration as is exemplified in the next section.

## TOWARDS THE PRACTICE OF CORPUS SHARING

### Socializing the Corpus through a Data Session

When a group of researchers is appropriating a corpus, the processes at work can be greatly facilitated through the organization of collective work sessions. Emanating from the field of conversation analysis, the format of the data session corresponds to an academic practice carried out by a group of researchers during which data recordings are projected so that the session’s participants can explore data collectively (Caria, 2010). As ten Have (1999) explains, what is at stake with a data session is not what any individual researcher can make of a data set *per se*, but “rather of sharable and shared understandings” (p. 123). Ten Have proceeds with the metaphor of the data session as a playground, both as an individual effort of

grounding proposed analyses in the actual data (the rules of the game) and as a collective endeavor to reach a shared understanding of a corpus (the aim of the game).





Caria (2010) details the progress of a data session through four steps:


1. The data session leader projects an extract or a collection of extracts to his fellow researchers along with transcriptions of the verbal output. The projection is repeated as much as needed.
2. Participants take a few minutes to gather their ideas and annotate their transcriptions.
3. Participants take turns to formulate their observations, show their interest in a given phenomenon, and discuss methodological or theoretical aspects.
4. Participants are then invited to react to what their colleagues have said, offer alternative interpretations, and propose additional comments or observations.

For the ISMAEL project, a two-day workshop was organized in order to familiarize the various researchers with the corpus and the characteristics of the pedagogical situation. The data session was given a central place in the workshop because such an academic format was deemed important to facilitate the discovery of the corpus and encourage researchers from different backgrounds and experience to work collectively. Indeed, the original members of the ISMAEL group included researchers specialized in language education, teacher education, and conversation analysis. Most had worked in classroom language-teaching situations but had little or no experience with the specificities of synchronous online teaching. Before the workshop, four extracts from the ISMAEL corpus had been selected and sent to the participants because they were thought to provide opportunities to identify relevant characteristics of the pedagogical situation.

To illustrate the process of appropriation that can take place around a corpus, I propose a multimodal analysis of one extract (see [Clip 3](#)) from the data session which brought together 14 participants (including the author of the present paper who was also the data session leader). All the researchers were seated around an oval table, which allowed them to see one another, and they took turns by raising their hands. A 360-degree camera placed in the middle of the table and a static camera were used to film the interactions. The captured interactions were transcribed according to conventions (see [Appendix B](#)) which gave an account of multimodality. [Transcript 1](#) works as a *mise en abîme*: it starts with the projection of the extract from the ISMAEL corpus already presented earlier (see [Clip 1](#)) with a trainee teacher and her two students. It is then followed by the interpretations proposed by two researchers from the ISMAEL project. The transcript of this part of the data session helps to understand what is at stake when a group of researchers participate in the collective exploration of multimodal data. Although it lasts only eight seconds, the clip that is projected (see [Line 1](#)) is sufficient to trigger rich interactions, which I now analyze.

## Transcript 1. Extract from the data session

1		<p>Adèle essayez de dire toutes les deux c'est votre [projet]  <i>try to talk both of you it's your project</i></p> <p>Catriona [hum:]</p> <p>Adèle et moi je vous note\ ++ moi je suis le manager je  [je vous juge]  <i>and I'm grading you ++ I am the manager I'm judging you</i></p> <p>Alannah [okay]</p> <p>Catriona [hum:]</p>
2		<p>NG: on peut peut-être arrêter dès maintenant je pense c'est le  <i>we can stop right there I think it's the</i></p> <p>(glances around to see which of the participants wants to take the floor first)</p>
		silence and laughter
3.1		<p>CW: alors moi je trouve intéressant le geste autour de <u>je vous juge</u>/  suivi d'un rire  <i>so I find interesting the gesture for I'm judging you/ followed by a laugh</i></p>
3.2		<p>CW: est-ce que c'est pour essayer de faire passer le message  d'évaluateur plus facilement/  <i>is it to try to get the evaluative message across more easily</i></p>
4.1		<p>FC: euh passer le message ++ il y a deux + il y a probablement +  peut-être pas aller trop trop loin pour la signification du rire faut  pas aller trop vite  <i>uh to get the message across ++ there are two + there are probably + maybe not go too too far for the meaning of the laugh let's not go too quickly</i></p>
4.2		<p>FC: moi j'y verrais deux choses + un elle atténue + c'est un signal  d'atténuation <u>puisque'elle s'est mise en position haute</u> je suis le  manager/  <i>I would see two things + one she mitigates + it is an attenuation signal since she puts herself in an authoritative position I am the manager/</i></p>

4.3		<p>FC: et l'autre hypothèse peut-être pour le rire c'est que c'est fictionnel/ + et on sait ça enfin moi pour l'avoir beaucoup regardé que quand on change de régime + de régime fictionnel au régime sérieux/ + catégories de Bange/ + euh y a le rire qui montre + je suis le manager/ + c'est pas vrai/ + <u>et on rit</u> + oui oui c'est une marque euh + il a donc peut-être plusieurs euh + significations le rire là</p> <p><i>and the other hypothesis maybe for the laugh is that it's fictional/ + and we know that like for me having studied it a lot that when you change mode from a fictional mode to a serious mode + Bange's categories/ + uh there's the laugh that shows + I'm the manager/ + it's not true/ + <u>and we laugh</u> + yes yes it's a mark uh + it therefore maybe has many uh + significations that laugh there</i></p>
-----	---	--

First, it can be noted that the two researchers, Ciara Wigham (CW) and Francine Cicurel (FC), weave the most salient features of the interaction into their own interpretations by repeating both parts of the teacher's verbal message (*je vous juge, je suis le manager*) and some co-verbal aspects (the teacher's raised finger [Line 3.1], her smile [Line 4.3]). Corpus exploration seems to require researchers to borrow from the studied interaction and subsequently cite relevant passages as if there were a need to inscribe the participants' behaviors and words into their own bodies to appropriate them fully.

Besides, the appropriation process also entails several operations:

- Identifying salient features (Line 3.1), that is, the conjunction of the teacher's raised finger and the quasi-synchronous production of a smile creating an oxymoronic message (*I'm judging your oral production while I maintain a benevolent attitude*).
- Using some established notions, such as softeners (Line 4.2), to explain how the teacher is trying to reduce the asymmetry between her and her learners.
- Using some references, as in Line 4.3, from the literature (*Bange*) and past research (*having studied it a lot*) to embed the interpretations that are proposed within a scientific tradition and thus augment their legitimacy.

This collective exploration of the data proved crucial as it allowed participants who were not familiar with technology-mediated teaching to understand the didactic contract—that of online pedagogical conversation—which Guichon & Drissi (2008) define as a genre hybridizing classroom and social talk. Selecting such an episode for the data session and scrutinizing it collectively thus helped the group members to identify the specific type of communication and didactic contract that shaped the exchanges included in the corpus. The study of the whole data session also revealed that working collectively on some extracts was key to familiarizing the researchers with the participants (e.g., the trainees' relative online teaching inexperience and the learners' levels). It also gave insiders opportunities to provide outsiders with essential contextual explanations and point them to complementary data that could be used to enrich their understanding.

### Multimodal Representation and Dissemination




Another outcome of socializing a corpus consists of collectively devising the different ways of representing the collected data in the transcripts that accompany future textual interpretations. Collective reflection and actual trials at making transcripts can prove invaluable to raise participants' awareness of the multimodal aspects of the corpus and of the representational challenges posed by multimodal research. These include three issues:

- what to include in the transcript and the necessary tradeoff between rendering all the aspects of the data and keeping the transcript readable;
- how to represent the interaction between the multiple modes; and
- how to represent the orchestration of semiotic resources by interactants.

Ten Have (1999) puts forward the meditational function of transcripts and their contribution “to highlight specific phenomena and create a ‘shared focus’ among audience and analyst” (p. 33). In the making of a multimodal transcript, choices that are made by researchers (e.g., how to name the actors, whether or not to translate interactions, etc.) contribute to proposing one representation of the data. To retain the quality of online pedagogical interactions, transcripts which include screen casts facilitate the emergence of some dimensions that might otherwise be overlooked.

In the first stage of our project, it was decided that the transcripts would not need to be homogenized so as not to curb the researchers’ creativity even if two main representations (one akin to a musical score as in [Transcript 1](#) and the second organized as a vertical strip as in the [Transcript 2](#)) emerged from discussions.

[Transcript 2](#). Representation of the oxymoronic “je vous juge”

(1)	(2)	(3)
		
Adèle essayez de dire toutes les deux c'est <u>votre</u> [projet Alannah [okay] Catriona [hum:]	Catriona [hum:] Adèle et] moi j[e vous note\ moi je suis le manager je [je <u>vous juge</u> ]	(laughter) Alannah [okay] Catriona [hum:]

When put together collectively, transcripts work as intermediary objects that force participants to find common ground to analyze a given linguistic phenomenon. Thus, discussing the best ways to describe and represent an online interaction is a means of aligning researchers coming from different perspectives with the same scientific purpose and ultimately co-constructing a shared representation while building a community around the same object.

Concerning dissemination, the CALL community would benefit from rethinking the way it communicates its results and analyses as is advocated by the proponents of digital humanities (e.g., Hayles, 2012). In line with this trend, I contend that published texts could be presented more systematically and be accompanied by multimodal documents in order to render the specificities of a given teaching or learning situation more holistically. To do so, such journals as *Language Learning & Technology* are adapted venues since they provide the (still rare) opportunity of publishing texts accompanied by relevant media. Such online multimodal publications could enable readers to “analyze the data themselves [and] test the analytical procedures which the author has followed and the validity of his/her analysis and claims” as is advocated by Seedhouse (2005, p. 254). A multimodal presentation of research results would not only contribute to improving their “intellectual accountability” (Markee & Stansell, 2007, p. 37) but also help build “communities of scholarly practice” (p. 36) among researchers by giving them a remote access to data that can then be shared and analyzed further (González-Lloret, 2015).

Thanks to “the emotional affordances of these rich media” (Lemke, 2013, p. 68), edited video clips can be used to share with readers what research actually feels like, an endeavor all the more important when research is not only targeted at other scholars but aimed to reach out to teachers. As ethnographer Pink (2006) pointed out, images can be as meaningful as texts to give an account of a piece of research, and she insists that “images may not necessarily be *the* main research method or topic, but through their relation to the sensory, material, and discursive elements of the research images and visual knowledge will *become* of interest” (p. 5). When relevant, associating image and text in research dissemination is of prime importance for a semiotic approach. Online publication should contribute, even more than it does today, to communicating the outcomes of CALL research.

Finally, reaching out to an audience beyond the sole research community through online versions of research outcomes could help fuel the reflection of current and future language teachers. Designing training resources in dialogue with research and enhancing their shareability with the multimodal means that are now available could thus continue reframing CALL research into a design-based endeavor whereby “theoretical issues get tested in the design of implementations, and implementations are loci of theoretical reflection and elaboration” (Burdick et al., 2012, p. 13). The edited book by Guichon and Tellier (2017) that is one of the outcomes of the ISMAEL project is accompanied by an [online repository](#) where clips from the corpus, definitions, and relevant resources have also been shared with the community of language teachers.

## CONCLUSION

The main aim of this article is to show that a corpus-based approach used to gather a large dataset of online interactions could provide a pertinent methodological framework when carrying out research focusing on online teachers’ professional development. Such an approach is a fruitful avenue for CALL research as it provides researchers with the opportunity to scrutinize the intricate working of a mediated situation that can be reconstituted thanks to the traces left by the participants in that situation. A corpus-based approach therefore requires that researchers become intimate with these traces and appropriate them through different operations of collection, transformation, analysis, and re-presentation that have been described in this article. This comprehensive process of appropriation is eminently ethical in that it turns the linguistic practice of corpus analysis into a relentless effort to reduce the strangeness of the traces produced by actors in situations where the researchers (and the future readers of the research) were not involved, without reducing the complexity of the situation itself.

I have proposed the concept of corpus shareability and have used some aspects of the ISMAEL project to demonstrate that attention to ethics, data accessibility, and corpus socialization could all facilitate corpus sharing. I have claimed that corpus sharing can contribute vastly to the field of CALL by enhancing its scientific robustness as it favors a multidisciplinary, systematic, and in-depth analysis of multimodal data. Corpus sharing could also be extended beyond researchers who share the same epistemological stance. It is imaginable and even desirable that once a qualitative approach has been used, such as the one described in this paper, a corpus linguistics-based methodological approach could then be used to investigate the same data through another lens and thus reveal aspects that are overlooked by a qualitative approach, providing “a utility for [corpus] re-usability” (Knight et al., 2009, p. 9).

Finally, this article devoted to methodology has endeavored to underline the human aspects of research. This emphasis on corpus creation, corpus sharing, and concrete illustrations of a research community involved in corpus-based research stems from a resolute wish to demonstrate that methodology is not a linear, recipe-like and disembodied process. On the contrary, it is caught in a web of sometimes conflicting scientific traditions, ethical constraints, technological affordances, and human activity and affect. While I do not suggest that research should become self-centered and unduly anecdotal, I believe that it might be important to provide the community of researchers—especially those in training—with glimpses of research in the making (Latour, 2005) and unveil some of the activity that takes place before



dissemination.

### APPENDIX A. Extract form the Consent Form Used for the Project

Participation in this research project is voluntary and your decision to take part or not will have no effect on your grades. If you decide not to take part, the data you authored will not be included in the corpus.

Please complete the following (Circle Yes or No for each question)

*I have read the Plain Language Statement (or had it read to me) Yes No*

*I understand the information provided Yes No*

*I have had an opportunity to ask questions and discuss this study Yes No*

*I have received satisfactory answers to all my questions Yes No*

*I am satisfied that arrangements have been made to protect my anonymity as far as possible Yes No*

*I am aware that the confidentiality of the information I provide is subject to legal limitations Yes No*

*I give permission for my anonymised data to be included in the corpus. Yes No*

*I am aware that the corpus may be shared with other researchers Yes No*

I give permission for extracts of the recordings where I appear to be shown at conferences and teacher training sessions:

*With my face in clear but with all other identity markers masked out Yes No*

*With my face blurred and all other identity markers masked out Yes No*

I give permission for snapshots of the recordings where I appear to be published in academic articles

*With my face in clear but with all other identity markers masked out Yes No*

*With my face blurred and all other identity markers masked out Yes No*

*I give permission for selected anonymised quotes from my data to be used in academic publications Yes No*

*I would be willing to take part in interviews/focus groups at a future date Yes No*

I have read and understood the information in this form. My questions and concerns have been answered by the researchers, and I have a copy of this consent form. Therefore, I consent / do not consent (*strikethrough as appropriate*) to take part in this research project under the conditions specified above.

Participants Signature: \_\_\_\_\_

Name in Block Capitals: \_\_\_\_\_

Witness: \_\_\_\_\_

Date: \_\_\_\_\_

### APPENDIX B. Transcription Conventions

[ ] Overlapping talk

/\ Rising or falling intonation

◦◦ Lower voice

::: Extension of the sound or the syllable it follows

p`tit Elision

trouv-	Truncation of a word
xxx	Incomprehensible syllabe
=	Latching
()	Uncertain transcription
(( ))	Comments
&	Turn of the same speaker interrupted by an overlap
(.)	Micro-pause
(0.6)	Timed pause
<u>it was good</u>	Underlined segments to show the beginning and the end of a gesture

---

## ACKNOWLEDGEMENTS

The ISMAEL project was supported by the ASLAN project (ANR-10-LABX-0081) of Université de Lyon, within the program “Investissements d’Avenir” (ANR-11-IDEX-0007) of the French government operated by the National Research Agency.

I am indebted to all the participants in Dublin and in Lyon and all the researchers who have taken part in the ISMAEL project and have added value to the article: Brahim Azaoui, Françoise Blin, Lucile Cadet, Francine Cicurel, Cathy Cohen, Isabel Colón de Carvajal, Thomas Debay, Charlotte Dejean, Eugénie Duthoit, Marie-Josée Hamel, Benjamin Holt, Justine Lascar, Cédric Sarré, Marion Tellier, Sylvie Thouésny, Daniel Valéro, Julie Vidal, and Ciara Wigham.

---

## ABOUT THE AUTHOR

Nicolas Guichon is a professor in language sciences at the Université de Lyon and belongs to the ICAR (*Interactions, Corpus, Apprentissages, Représentations*) research team. His research interests include teacher education in CALL, the study of online interaction, and digital literacies.

**E-mail:** [nicolas.guichon@univ-lyon2.fr](mailto:nicolas.guichon@univ-lyon2.fr)

---

## REFERENCES

- Adolphs, S., & Varter, R. (2013). *Spoken corpus linguistics. From monomodal to multimodal*. New York, NY: Routledge.
- British Psychological Society. (2013). *Ethics guidelines for internet-mediated research*. Retrieved from: <http://www.bps.org.uk/system/files/Public%20files/inf206-guidelines-for-internet-mediated-research.pdf>
- Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., & Schnapp, J. (2012). *Digital humanities*. Cambridge, MA: MIT Press.
- Cappeau, P., & Gadet F. (2007). L’exploitation sociolinguistique des grands corpus. Maître-mot et pierre philosophale. *Revue Française de Linguistique Appliquée*, 121(1), 99–110.
- Caria, A. (2010). L’information ethnographique en analyse de conversation: Le cas de la data session. *Cahiers de Praxématique*, 54–55, 189–206.

- Chanier, T., & Ciekanski, M. (2010). Utilité du partage des corpus pour l'analyse des interactions en ligne en situation d'apprentissage: Un exemple d'approche méthodologique autour d'une base de corpus d'apprentissage. *Alsic*, 13. doi: 10.4000/alsic.1666 Retrieved from: <http://alsic.revues.org/1666>
- Chanier, T., & Wigham, C. R. (2016). A scientific methodology for researching CALL interaction data: Multimodal LEarning and TEaching Corpora. In M.-J. Hamel, & C. Caws, (Eds.), *Learner computer interactions: New insights on CALL theories and applications*. (pp. 215–240). New York, NY: John Benjamins.
- Develotte, C., Guichon, N., & Vincent, C. (2010). The use of the webcam for teaching a foreign language in a desktop videoconferencing environment. *ReCALL*, 23(3), 293–312.
- Dooly, M. (2015). It takes research to build a community: Ongoing challenges for scholars in digitally-supported communicative language teaching. *CALICO Journal*, 32(1), 172–194.
- Dybkjær, L., & Ole Bernsen, N. (2004). *Recommendations for natural interactivity and multimodal annotation schemes*. Paper presented at the LREC 2004 Workshop on Multimodal Corpora, Lisbon, Portugal.
- Flewitt, R., Hampel, R., Hauck, M., & Lancaster L. (2009). What are multimodal transcription and data? In C. Jewitt, (Ed.), *The Routledge handbook of multimodal analysis*. (pp. 40–53). London, UK: Routledge.
- González-Lloret, M. (2015). Conversation analysis in computer-assisted language learning. *CALICO Journal*, 32(3), 569–594.
- Granger, S., Gilquin, G., & Meunier, F. (2015). *The Cambridge handbook of learner corpus research*. Cambridge, UK: Cambridge University Press.
- Guichon, N. (2013). Une approche sémio-didactique de l'activité de l'enseignant de langue en ligne: Réflexions méthodologiques. *Education & Didactique*, 7(1), 101–115. Retrieved from <https://educationdidactique.revues.org/1679>
- Guichon, N., & Cohen, C. (2014). The impact of the webcam on an online L2 interaction. *Canadian Modern Language Review*, 70(3), 331–354.
- Guichon, N., & Drissi, S. (2008). Tutorat de langue par visioconférence: Comment former aux régulations pédagogiques. *Les Cahiers de l'ACEDLE*, 5(1), 185–217.
- Guichon, N., & Tellier, M. (2017). *Enseigner l'oral en ligne. Une approche multimodale*. Paris, France: Didier.
- Guichon, N., Bétrancourt, M., & Prié, Y. (2012). Managing written and oral negative feedback in a synchronous online teaching situation. *Computer Assisted Language Learning*, 25(2), 181–197.
- Guichon, N., Blin, F., Wigham, C. R., & Thouësny, S. (2014). *ISMAEL LEarning and TEaching Corpus*. Lyon, France: Laboratoire Interactions, Corpus, Apprentissages & Représentations.
- Hampel, R., & Stickler, U. (2012). The use of videoconferencing to support multimodal interaction in an online language classroom. *ReCALL*, 24(2), 116–137.
- Hayles, K. N. (2012). How we think: Transforming power and digital technologies. In D. M. Barry (Ed.), *Understanding digital humanities* (pp. 42–66). New York, NY: Palgrave Macmillan.
- Jewitt, C. (2009). *The Routledge handbook of multimodal analysis*. London, UK: Routledge.
- Jones, R. H. (2004). The problem of context in computer-mediated communication. In P. Levine & R. Scollon, (Eds.), *Discourse and technology—Multimodal discourse analysis* (pp. 20–33). Washington, DC: Georgetown University Press.

- Kern, R. (2014). Technology as Pharmakon: The promise and perils of the Internet for foreign language education. *Modern Language Journal*, 98(1), 340–357.
- Knight, D., Evans, D., Carter, R., Adolphs, S. (2009). HeadTalk, HandTalk and the corpus: Towards a framework for multi-modal, multi-media corpus development. *Corpora*, 4(1), 1–32.
- Latour, B. (2005). *La science en action. Introduction à la sociologie des sciences*. Paris, France: La Découverte.
- Lemke, J. L. (2013). Thinking about feeling: Affect across literacies and lives. In O. Erstad & J. Sefton-Green (Eds.), *Identity, community, and learning lives in the digital age* (pp. 57–69). Cambridge, UK: Cambridge University Press.
- Markee, N., & Stansell, J. (2007). Using electronic publishing as a resource for increasing empirical and interpretive accountability in conversation analysis. *Annual Review of Applied Linguistics*, 27, 24–44.
- Norris, S. (2004). *Analyzing multimodal interaction—A methodological framework*. New York, NY: Routledge.
- Pink, S. (2006). *Visual ethnography. Images, media, and representation in research*. London, UK: Sage Publications.
- Satar, M. (2013). Multimodal language learner interactions via desktop videoconferencing within a framework of social presence: Gaze. *ReCALL*, 25(1), 122–142.
- Seedhouse, P. (2005). Conversation analysis as research methodology. In K. Richards & P. Seedhouse (Eds.), *Applying conversation analysis* (pp. 251–266). New York, NY: Palgrave MacMillan.
- Sindoni, M. G. (2013). *Spoken and written discourse in online interactions*. New York, NY: Routledge.
- Sloetjes, H., & Wittenburg, P. (2008). *Annotation by category—ELAN and ISO DCR*. Paper presented at the 6th International Conference on Language Resources and Evaluation, Marrakesh, Morocco. Retrieved from [http://www.lrec-conf.org/proceedings/lrec2008/pdf/208\\_paper.pdf](http://www.lrec-conf.org/proceedings/lrec2008/pdf/208_paper.pdf)
- ten Have, P. (1999). *Doing conversation analysis*. London, UK: Sage Publications.
- Vinck, D. (1999). *Ingénieurs au quotidien. Ethnographie de l'activité de conception et d'innovation*. Grenoble, France: PUG.
- Yamada, M., & Akahori, K. (2009). Awareness and performance through self- and partner's image in videoconferencing. *CALICO Journal*, 27(1), 1–25.
- Yanguas, I. (2010). Oral computer-mediated interaction between L2 learners: It's about time! *Language Learning & Technology*, 14(3), 72–93. Retrieved from: <http://llt.msu.edu/issues/october2010/yanguas.pdf>