



**HAL**  
open science

## A comparative study of lexical word search in an audioconferencing and a videoconferencing condition

Cathy Cohen, Ciara R. Wigham

### ► To cite this version:

Cathy Cohen, Ciara R. Wigham. A comparative study of lexical word search in an audioconferencing and a videoconferencing condition. *Computer Assisted Language Learning*, 2018. halshs-01860136

**HAL Id: halshs-01860136**

**<https://shs.hal.science/halshs-01860136>**

Submitted on 23 Aug 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**This is the authors' copy of an accepted manuscript to be published in *Computer Assisted Language Learning* (Routledge, Taylor & Francis Group)**

**A comparative study of lexical word search in an audioconferencing and a videoconferencing condition**

**Cathy Cohen<sup>1</sup>, Ciara R. Wigham<sup>2</sup>**

*<sup>1</sup>Laboratoire Interactions, Corpus, Apprentissages & Représentations (ICAR) – Université Lyon 1*

*<sup>2</sup>Laboratoire de Recherche sur le Langage (LRL) – Université Clermont Auvergne*

This study on online L2 interactions compares lexical word search between an audioconferencing and a videoconferencing condition. Nine upper-intermediate learners of English describe a previously unseen photograph in either the videoconferencing or the audioconferencing condition. A semantic feature analysis is adopted to compare their interactions. To evaluate the contribution of visual and verbal modes, a quantitative analysis examines the distribution of the referential properties of one target lexical item: *tunnel earring*. It suggests that pushed output produced in the videoconferencing condition is lexically richer. Then, in view of these results, focusing on two learners, one from the audioconferencing condition and one from the videoconferencing condition, a fine-grained multimodal analysis of the qualitative features of gestures and speech complements the quantitative results. It demonstrates how the videoconferencing condition allows the learner to embody salient physical referential properties of the lexical item, before transferring the referential information to the verbal mode, to produce a semantically rich description. The study will interest researchers working on multimodality and L2 teachers deciding between videoconferencing and audioconferencing as pedagogical options.

Keywords: computer-mediated communication; distance learning; speaking skills; multimodality; online teaching and learning

## Introduction

Online language teaching is integrating features that allow greater multimodality and synchronicity. Synchronicity is considered to enhance online pedagogical interactions (Hrastinski, 2008; Levy & Stockwell, 2006). Although the contributions of multimodality to language learning have been studied within audioconferencing or videoconferencing environments, research comparing teaching situations which incorporate a broader or narrower range of channels is scarce.

Indeed, initial studies focused on how synchronicity and multimodality within audioconferencing environments may enhance learners' oral participation, speaking skills and collaboration (Hampel & Hauck, 2004; Ciekanski & Chanier, 2008; Vetter & Chanier, 2006). More recent studies have investigated the contributions of the webcam in videoconferencing environments and the ways in which the interlocutor's image, that gives access to communicative resources including gestures, facial expressions, body movements and gaze, may contribute to more active communication and better mutual understanding. Such studies have explored analysis units including social presence (e.g., Guichon & Cohen, 2014; Satar, 2013), lexical explanations (e.g., Holt & Tellier, 2017; Wigham, 2017), word search (Cappellini, 2013; Nicolaev, 2012), teacher semio-pedagogical competence (e.g., Guichon & Wigham, 2016; Kozar, 2016) and task design (e.g., Hauck & Youngs, 2008).

Audioconferencing and videoconferencing environments which offer synchronicity both offer new pedagogical possibilities for developing oral language skills in interaction (Guichon & Tellier, 2017). Although online language teaching institutions often must decide as to whether to include audioconferencing or videoconferencing in their online language learning programmes, few research studies have compared these conditions (Guichon & Cohen, 2014). In addition, such studies are often based on learners' perceptions (Rosell-Aguilar, 2007) and empirical studies frequently report divergent findings (Guichon & Cohen, 2014; Yamada & Akahori, 2007; 2009; Yanguas, 2010).

The current study therefore compares the analysis unit of lexical word search between a videoconferencing and an audioconferencing condition. Within the Second Language Acquisition (SLA) field, studies have examined native and non-native speakers' word search practices as interactional phenomena (Brouwer, 2003; Jung 2004; Kurhila, 2006). Word searches are considered "crucial moments in the

learner's acquisition of target language structure" (Hammarberg, 1998, p. 178). Focusing on teacher-learner interactions using Skype, either with the webcam on (videoconferencing), or off (audioconferencing), our study compares the verbal content of learners' lexical explanations; explores to what extent multimodal semiotic resources enhance and complement oral language output in the videoconferencing condition; and investigates how close the learners approach the conventional meaning of the target lexical item in their descriptions in each condition.

The current study utilises data from a previous quantitative study (Guichon & Cohen, 2014) that included a comparative overview of the number and duration of word search episodes across the two conditions. Findings revealed no significant differences for these two variables. However, the study did not explore how verbal and visual resources were used during word search, nor the affordances of the webcam image. The current follow-up study therefore re-examines a subset of the data by focusing on one lexical item: *tunnel earring*. A semantic feature approach allows an initial quantitative overview of nine participants. This approach examines how the different semiotic resources are used between the two conditions. First we compare the verbal mode across the two conditions, before adding the visual mode to the videoconferencing condition to explore the contribution of gestures. Next we conduct a fine-grained qualitative analysis on two learners, one from each condition, to further explore one result from the quantitative overview; that the verbal mode in the videoconferencing condition is richer in semantic information.

The paper proceeds as follows: a literature review firstly addresses the notion of 'word search', the communicative functions of gestures in studies on non-pedagogical native-speaker interactions and in studies on second language learners, and an overview of previous studies comparing the communicative affordances of audioconferencing and videoconferencing. After detailing our research questions, we explain the research design and analysis methodology. Findings from the quantitative overview and subsequent qualitative analysis are then reported before being discussed with regard to the research questions and previous literature.

## **Word Search**

Word searches are episodes when a speaker "displays trouble with the production of an item in an ongoing turn at talk" (Brouwer, 2003, p. 535) rather than providing the

next relevant turn. The current paper focuses on a lexical search where a speaker wishes to label a concept but does not have, or cannot recall or retrieve, the necessary resources (Kasper & Kellerman, 1997). Word search has been studied from a multimodal perspective in both L1 and L2 interactions, as discussed below. In the verbal mode, word search is observed in private speech, which is typically defined as “speech addressed to the self (not to others) for the purpose of self-regulation (rather than communication)” (Diaz & Berk, 1992, p. 62). Word search is marked by sound stretches or hesitation signals accompanied by falling intonation or produced at a lower volume, to signal that a participant is engaging in a word search (Brouwer, 2003; Goodwin & Goodwin, 1986; Hauser, 2003; Kurhila, 2006). Studies have also described how speakers may communicate they are entering a word search by displaying relative unavailability through withdrawn gaze and production of a thinking face (Goodwin & Goodwin, 1986) that may help the speaker hold the floor (Carroll, 2005).

Several studies have examined how participants attempt to solve word searches. In the verbal mode, strategies include using loan words (Jung, 2004; Kurhila, 2006), foreignising words from other languages as glosses (Kurhila, 2006) and using words which are related semantically to describe the target item by employing synonymic, metonymic, antonymic or superordinate relations (Kurhila, 2006). Negation is often part of this latter strategy to help delimit and define the semantic field of the target item. The visual mode allows a speaker to make apparent changing progress in the search: facial expressions reveal consecutive attempts to recover the target item (Goodwin & Goodwin, 1986).

If recipient participation is involved in the word search resolution, specific interactional work must be accomplished. Strategies include producing an account for not providing the item, specifically addressing the other speaker in the word search marker (Brouwer, 2003) and, in the visual mode, using gaze-change (Carroll, 2005).

Word searches are considered complete when the participants reach intersubjective understanding. Jung (2004) has demonstrated that word search completion may lead to repetition of the target lexical item. Brouwer (2003) considers repetition of this type as a local demonstration of learning. However, a word search may also be abandoned if, despite a multitude of efforts to establish mutual

understanding (Egbert, Niebecker, & Rezzara, 2004), the participants fail to reach subjective understanding.

Although word search has been studied in face-to-face SLA contexts, to the best of the authors' knowledge, studies focusing on L2 word-search in CALL remain rare. Nicolaev (2012) explored metalinguistic sequences during task-based interactions via videoconferencing between trainee-teachers and learners of French. The study illustrated how learners invited participation through explicit or implicit verbal requests, often in the L1, and visually through deictic (i.e., pointing) or iconic gestures (i.e., gestures providing a visual image closely related to the semantic content of the co-occurring speech). In the corpus, the target item was principally proffered by the tutors through quasi-simultaneous use of audio and text chat modalities.

Cappellini (2013) built on Nicolaev's study on word search by examining how learners in Chinese-French teletandem interactions described the sought lexical item accompanied by iconic gestures. Code-switching played a central role in signalling word search onset, often combining L2 audio interaction and L1 use in the text chat. Nicolaev and Cappellini's work has emphasised the importance of gestures in word search. Thus, we now turn to their possible communicative function.

## **Communicative function of gestures**

### ***Communicative function of gestures in studies on non-pedagogical native-speaker interactions***

One area in gesture research has investigated native-speakers in non-pedagogical interactions completing different communication tasks (e.g., describing images, objects; giving directions), while varying mutual visibility between speakers and comparing gesturing across conditions (e.g., Bavelas, Gerwing, Sutton, & Prevost, 2008; Cohen & Harrison, 1973). Participants are therefore communicating either face-to-face, or without mutual visibility (on the telephone, for instance). These studies are referred to as *visibility studies*.

Although the aims of the current study differ from these native-speaker visibility studies, we draw upon their methodologies and analyses to explore data from our online L2 visibility study. Indeed, our interest lies in assessing whether an audioconferencing or videoconferencing condition is more facilitative of

communication during word search in online pedagogical interactions. Therefore, as in the native-speaker visibility studies discussed below, the communicative functions of the available semiotic resources in each condition are analysed: the audioconferencing condition offers solely the verbal mode, while visual resources are available in the videoconferencing condition, enabling speakers to create “integrated messages” (Gerwing & Allison, 2011, p. 309).

Several native-speaker visibility studies have investigated how seeing one’s interlocutor influences the qualitative features of gestures, including size and shape (e.g., Alibali, Heath, & Myers, 2001; Bavelas et al., 2008). In the latter study, participants described a picture of an elaborate dress to an interlocutor. Results showed that participants in the face-to-face condition often produced life-sized gestures positioned around their own body to show the location or scale of certain features. Alibali et al. (2001) argued that gestures in face-to-face interactions were generally more elaborate and iconic than gestures employed when interlocutors could not see one another.

Certain studies (Bangerter, 2004; Bavelas et al., 2008; Emmorey & Casey, 2001; Gerwin & Allison, 2011; Melinger & Levelt, 2004) have analysed how visibility impacts the speech-gesture relationship. Findings showed that in face-to-face interactions, since gestures had unique communicative functions, some were redundant with speech while others carried information that was absent from the verbal mode. By contrast, without mutual visibility, interactants’ words often carried a heavier informational load, compensating for the lack of co-verbal resources.

Using data from Bavelas et al. (2008), Gerwin and Allison (2011) applied a semantic feature analysis (see also Beattie & Shovelton, 1999) to explore the semantic contribution of gestures and speech in the face-to-face condition compared to the condition without mutual visibility. In a semantic feature approach, categories of information on the target lexical item are first identified in participants’ speech. Then an assessment is made as to whether words or gestures (or both) contribute information about each category. So having identified certain salient dress characteristics, Gerwin and Allison (2011) assessed how information in different semantic categories was distributed through gestures, speech, or both. They hypothesised that if gestures had a communicative function, the relative semantic distribution of information between gestures and speech would change, depending on whether interlocutors could see one another. Gerwin and Allison (2011) showed

that gestures in face-to-face conversations conveyed more information than speech and that the verbal mode in telephone interactions carried more information than in the face-to-face condition. A growing number of studies have investigated the communicative functions of gestures in L2 learners' speech as will now be discussed.

### ***Communicative function of gestures in studies on second language learners***

Several studies have compared the use of gestures by individuals speaking their L1 and L2 (Gregersen, Olivares-Cuhat, & Storm, 2009; Gullberg, 2009; Stam, 2006a). Overall findings indicated that gesture rate was more frequent when speaking an L2 and higher in lower level learners (Faraco & Kida, 1998; Gullberg, 1998; Stam, 2006b). Gullberg (1998) investigated how learners used gestures when dealing with difficulties encountered while speaking. She found that gestures did not necessarily replace speech but accompanied it, serving various functions, such as flagging on-going word searches or eliciting speech.

Compared to native-speaker studies, studies on L2 speakers have indicated a higher degree of redundancy between speech and gesture; in other words, the verbal and visual modes carry the same information. McCafferty (2002) explored how an intermediate learner used iconic gestures that conveyed information which was also present in his speech but did not add to what was conveyed verbally. He hypothesised that these iconic gestures had a cognitive, self-regulatory function: they supported the thinking process and reduced cognitive load, helping the learner plan his speech which consequently facilitated L2 production. Gullberg (2006) conducted a visibility study in which she provided evidence that representative redundant gestures persisted even when participants could not see one another, thus demonstrating their internal, self-regulatory function. Likewise, iconic gestures supported word searches (Negueruela & Lantolf, 2008). McCafferty (2004) suggested that increased pressure to communicate might have accounted for more frequent self-regulatory gestures in L2 communication, as learners not only had to decide what they wished to say, but also how to say it. Nevertheless, representational gestures may still have served a communicative function, helping learners make themselves understood "by drawing on the underlying mimetic properties of gestural imagery" (Stam & McCafferty, 2008, p. 15).

## **Studies comparing the communicative affordances of audioconferencing and videoconferencing**

Few studies have compared audioconferencing and videoconferencing in language-learning situations with the aim of assessing the psychological or communicative potential of seeing one's interlocutor. Rosell-Aguilar (2007) explored the perceptions of Spanish tutors with regard to an online audiographic and a face-to-face learning environment. He revealed that, while tutors found similarities between the conditions, the lack of access to paralinguistic cues, including facial expressions and gestures, in the audiographic condition prevented tutors from perceiving learners' reactions.

Yamada and Akahori (2007; 2009) explored the perceptions of students in learner-learner dyads when completing an explanation task in one of four conditions: (a) videoconferencing with both learners' images; (b) videoconferencing with only the partner's image; (c) videoconferencing with only the learner's image; and (d) audioconferencing only. In their 2007 study that examined perceived consciousness of social presence, language learning and learning objectives, as well as productive performance, they concluded that the interlocutors' image was most effective in promoting consciousness of presence. This conclusion was strengthened in their 2009 study which focused on an explanation task. They concluded that, first, communication was easier when both images were present but learners had more negative perceptions when their partner's image was absent. Second, videoconferencing impacted positively on learners' ability to understand their interlocutor. However, they claimed that restricting communication channels may have enhanced L2 learning by eliminating the possibility of using co-verbal devices.

Yanguas's (2010) study explored negotiation-of-meaning in a jigsaw task seeded with unknown lexical items. Learners were assigned to an audioconferencing, a videoconferencing or a face-to-face condition. Results showed that more linguistic resources were used in the audioconferencing group than in the two conditions which provided mutual visibility. Interactionists have argued that more pushed output should foster SLA (e.g., Smith, 2004). However, Yanguas emphasised that, although learners in the videoconferencing group produced fewer words, more lexical items were fully understood compared to the audioconferencing group: access to a range of semiotic resources in the videoconferencing condition gave learners a

more precise understanding of the target items, although they produced less language.

The current study uses a subset of data from a larger experimental study (Guichon & Cohen, 2014) that examined whether audioconferencing or videoconferencing was more facilitative of L2 communication. In their 2014 study, the researchers investigated how seeing one's interlocutor influenced interaction patterns (number of silences, overlaps, turn-taking); number and duration of word search episodes and learner perceptions of the interactions. They concluded that the webcam was not as critical to the interactions as anticipated, with few comparisons reaching statistical significance. Concerning number and duration of word search, they found no significant quantitative differences between the two conditions. However, their study was not designed to explore the use of verbal and visual resources during word search, nor the affordances of the webcam image.

Consequently, the current paper proposes a follow-up study that focuses on one lexical item: *tunnel earring* (a hollow, tube-shaped variety of body piercing jewellery) which led to frequent word search episodes during the interactions.

## Research questions

The present study was designed to answer two main research questions:

- (1) How is the semantic information about the *tunnel earring* communicated in the videoconferencing condition compared to the audioconferencing condition?
  - How does the verbal content of the lexical explanations differ?
  - What role do gestures play in communicating different semantic features?
- (2) Are there differences, between the two conditions, in how close the learners approach the conventional meaning of the target item?

To answer these questions, first, a quantitative analysis describes the distribution of the referential properties of *tunnel earring* in the two conditions. Then, by comparing the *tunnel earring* word search of one learner in the videoconferencing condition to one learner in the audioconferencing condition, a fine-grained multimodal qualitative analysis allows the quantitative results to be illustrated and explored in greater depth.

To the best of our knowledge, our study is novel in several ways. First, it compares word search between a videoconferencing and an audioconferencing condition. Secondly, this comparison is undertaken in a language learning situation. Finally, it adopts a semantic feature analysis to compare the word search episodes in a language learning situation.

## **Methodology**

For the purposes of the current study, a sub-set of data on nine participants was taken from data explored in Guichon & Cohen (2014). The 2014 study concerned 40 participants. Here, we therefore give a brief overview of how study-participants were divided between the two experimental conditions, before turning to the nine participants selected for the current study.

## ***Participants***

Nine participants were selected for the current study from the 40 participants who partook in a larger study reported in Guichon & Cohen (2014). The 40 B2-level learners (mean age = 20 years and 2 months; SD = 1.32; range = 17.4-24.6), all native or near-native French speakers, were second-year undergraduate students minoring in English at a French university. The study was external to their weekly language lessons and designed for research purposes. Participation was voluntary: no incentives were offered. All participants scored above 24 out of 40 on the Quick Placement Test (QPT, Oxford University Press and University of Cambridge Local Examinations Syndicate, 2001). They were divided equally between the two experimental conditions (videoconferencing and audioconferencing) and matched according to placement test score, age and sex.

### Table 1. Data extent

Nine participants, all French native speakers from France, were selected for the current study because each of them performed a word search around the target lexical item *tunnel earring* (see Column 5 of Table 1) whilst the remaining 31 participants did not. Five of those selected were in the videoconferencing condition (2 male, 3 female) and four were in the audioconferencing condition (2 male, 2 female).

Mann-Whitney tests revealed no group differences for QPT scores ( $U = 9$ ,  $z = -.249$ ,  $p = .905$ ,  $r = .083$ ) or age ( $U = 13$ ,  $z = .738$ ,  $p = .556$ ,  $r = .25$ ) for these nine participants.

The teacher had several years of experience teaching face-to-face English university courses. She regularly used online tools for personal communication, including the chosen tool, Skype, but had never taught online. As the current study focuses on learner word search and the teacher's role was simply to steer the conversation, this was not considered to constitute a problem.

All participants signed a consent form which explained how confidentiality would be ensured. Participants agreed to being recorded and filmed for research/teaching purposes. No information concerning the study's aims was given to the teacher or the learners in order not to influence the interactions.

### ***Research Design***

Learner data were recorded in a quiet University office, chosen deliberately so that the participants would not be disturbed and to allow high quality audio recordings. The teacher partook in the interactions from her home, in order to increase the ecology of the interactions. Indeed, for participants in the videoconferencing condition, if the teacher had appeared to be in a familiar setting (language classroom or University office), learners may have equated the situation with an assessment activity and, in turn, this might have had an impact on any potential anxiety and inhibited their oral contributions.

In Guichon & Cohen's 2014 study, learners were given four previously unseen photographs to describe and interpret to the teacher. The photographs were selected because each image contained potentially problematic lexical items (e.g., loudspeakers, tunnel earring, wheelchair) that might trigger word search episodes. The current study focuses on one of these photographs in which several young people are attending an outdoor concert. The central character is wearing a tunnel earring, the target lexical item in the current study.

Skype was chosen for the interactions because it is free to use. In the videoconferencing condition, the interlocutor saw the other participant's image full-screen and his/her own webcam image in a smaller window. In the audioconferencing condition, no images were present. Before meeting the previously unknown teacher for the online pedagogical interaction, learners were instructed not

to use the text chat in either condition and to avoid using French.

The teacher was asked to complete each interaction in around ten minutes<sup>1</sup>. She was also asked to act as an experimental confederate, behaving, each time, as a first-time participant. To facilitate comparison between interactions both within and across the two conditions, she was instructed to provide minimal feedback and was given a prepared script to be used as a prompt (see Appendix A). It was felt that such a script would help facilitate the comparison of interactions between the learners in the two conditions and aid the management of data processing. If learners provided words in French, the teacher was told to feign a lack of understanding. If they provided insufficient details, she incited them to elaborate, using the script's prompt questions. Although restricting the teacher's interactions may have reduced the number of co-verbal resources produced by the participants in general, this approach was adopted to ensure that the interactions were maximally comparable and to reduce variance in interactions due to differences in the interlocutor's behaviour. The teacher participated in around five interactions per session, alternating between experimental conditions to avoid developing fixed routines.

### ***Data collection and analysis***

The interactions were recorded using the screen-capture software Camtasia, chosen as the trial version was freely available and simple-to-use.

Data were analysed using an adaptation of the semantic feature approach (Beattie & Shovelton, 1999; Gerwin & Allison, 2011) to examine the qualitative features of gestures and speech within a quantitative paradigm.

Firstly, for each interaction, verbal transcriptions were completed using ELAN (Sloetjes & Wittenburg, 2008). Conventions are provided in Appendix B. ELAN was chosen because it allows "the close observation of dialogue as a lively synchrony of words, gestures, and faces" (Bavelas, Gerwing, & Healing, 2014, p. 127) and because its frame-by-frame viewing option facilitates systematic observation.

---

<sup>1</sup> This time period was chosen following a pilot study during which six students having the same profile finished the task in approximately ten minutes. The reasons for imposing an approximate time limit and having a script to guide each interaction was to facilitate the comparison of the performance of the participants and the teacher in the two experimental conditions and to make data-processing manageable given the initial number of participants in Guichon and Cohen's 2014 study.

Secondly, word-search episodes for *tunnel earring* were identified. This item was chosen because the authors felt it had distinguishing features which might lead to a range of representative gestures. As the item centres on a body part likely to fall into the webcam field, any gestures were likely to be visible to the interlocutor. The target item, thus, appeared a good foundation for an initial exploratory study that endeavoured to compare how semantic information was communicated in two conditions which differed in their inclusion of a visual mode. The word search episodes were isolated from the moment a learner expressed hesitation or uncertainty until the turn in which the teacher used a verbal reception marker (e.g., 'okay') or a transition to a new activity was observed. Sound stretches, hesitation markers, pauses or direct expression of word search were considered as marks of hesitation or uncertainty, as were off-screen gazes and lip thinning.

Thirdly, co-verbal acts used in word-search episodes were annotated for participants in the videoconferencing condition. These included gestures, head movements and changes in gaze direction.

### *Definitions of referential properties*

In the semantic feature approach adopted, ten referential properties of *tunnel earring* were determined by combining properties identified from the photo prompt and properties identified through an initial global exploration of the data. These were subdivided into properties relating to physical features (location, position in ear, shape, size, material), body modification (process, result) and other learner strategies (synonym, comparison, judgement). Appendix C describes each referential property.

### *Data coding*

For both conditions, data were coded individually by the authors using the list of ten referential properties of a *tunnel earring* to identify referential properties in the verbal mode for each interaction. For example, 'they put the size up' was coded as both 'size' and 'process' and left blank in the other eight categories. Inter-rater agreement, concerning the verbal data, equalled 98.4% (agreement on 945 of the 960 decisions concerning the 96 contributions of information). Coding disagreements were re-examined until 100% agreement was reached.

For the videoconferencing condition, both coders recorded decisions concerning the referential properties to which the visual mode contributed information using ELAN's frame-by-frame viewing. All co-verbal resources were analysed (51 contributions of information or 510 decisions as to whether the co-verbal resource depicted information related to each of the ten referential properties). Inter-rater agreement, concerning the visual mode, was 457 for the 510 decisions (89.6%). The interaction data were re-examined to reach full agreement.

We begin by conducting a quantitative analysis on data from the nine participants who performed a word search around the target lexical item *tunnel earring*. We compare the distribution of the referential properties of *tunnel earring* across the two conditions, first focusing only on the verbal mode, then adding the visual contributions in the videoconferencing condition. Next, in order to further account for the quantitative findings, we carry out a fine-grained multimodal qualitative analysis, following the methodological framework detailed in Norris (2004), in which one learner's word search in the videoconferencing condition is compared to one word search in the audioconferencing condition. These two learners were selected for analysis as we felt that, in terms of duration and content, their contributions were fairly representative of those of the other learners in the same condition who had performed a word search around this lexical item.

### *Statistical analyses*

The numerical data (from ELAN) were imported to SPSS (version 20) for analysis. Non-parametric tests were used in view of the small sample size and considering the assumption of normality was not met for several dependent variables (see Leech & Onwuegbuzie, 2002). Mean ranks are reported, following recommendations by Field (2013). Mann-Whitney tests were used for between-sample comparisons of the distribution of referential properties in the two conditions. Wilcoxon signed-rank tests were employed for within-sample comparisons of the distribution of referential properties in the verbal and visual modes of the videoconferencing condition. Two-tailed tests were used since the study is exploratory and no initial hypotheses are proposed. The alpha level was set at .05.

An effect size estimate ( $r$ ) is reported for each analysis following Fritz, Morris and Richler (2011). Benchmarks are adopted for  $r$  of: small ( $r = .25$ ), medium ( $r = .4$ ) and large ( $r = .6$ ), following Plonsky and Oswald's (2014) meta-analysis of the

calculation and use of effect sizes in the SLA field. Each analysis includes a reading for probability of superiority (*PS*) (see Fritz et al., 2011) to illustrate the size of the more important effects in a more meaningful, concrete way.

## **Analysis**

### ***Quantitative analysis***

Two sets of comparisons are conducted here. In the first set, we compare the verbal mode only between the audioconferencing and the videoconferencing conditions. In the second set, we compare the verbal plus visual modes in the videoconferencing condition to the verbal mode in the audioconferencing condition. For each set of comparisons, we start by comparing the overall distribution (*All properties*) of referential properties between the two conditions. Then we separate the data into the ten disaggregated properties and compare each to assess how semantic information was conveyed. Statistical results are reported in Appendix D.

#### ***Distribution of referential properties – verbal mode only***

The overall comparison (*All properties*) of the distribution in the verbal mode between the two conditions is not significant and the effect size is small (Appendix D-i). No statistically significant differences were established between the groups for any of the individual referential properties. However, for the physical characteristics, with the exception of *location*, the mean ranks for the videoconferencing condition are higher than for the audioconferencing condition, and they are identical for *size*. There is a large effect size for *shape*, with an equivalent *PS* of .8. In other words, if items were sampled randomly, one from each distribution, the number of references to *shape* in the verbal mode would be higher in the videoconferencing condition than the audioconferencing condition for 80% of comparisons. For the two body modification properties, the higher mean ranks and medium to large effect sizes suggest speakers' words in the videoconferencing condition are denser in information. The high *PS* scores confirm this, i.e., if sampled randomly, *process* and *result* would be richer in information in the verbal mode of the videoconferencing condition in 70% and 80% of comparisons respectively. Mean ranks are similar for the remaining referential properties, with small effect sizes and *PS* readings. So, a

high degree of overlap exists in the two conditions for *synonym*, *comparison* and *judgement*.

The high *PS* results and medium to high effect sizes reported for *shape*, *process* and *result* are intriguing and worthy of further investigation. Our subsequent qualitative analysis (see *Qualitative analysis* below) explores why the verbal mode in the videoconferencing condition is richer in semantic information relating to these categories.

#### *Distribution of referential properties – verbal and visual modes*

We now turn to the distribution of the referential properties between the two conditions, this time including all the contributions in the videoconferencing condition (verbal plus visual). Appendix D-ii provides a comparison of the overall distribution (*All properties*), followed by individual comparisons of the ten disaggregated properties.

The *All properties* comparison suggests that combining verbal and visual modes enables learners to convey more semantic information in the videoconferencing than the audioconferencing condition. This result is not significant, although there is a medium effect size.

For the comparisons of the ten disaggregated properties, results show that learners in the videoconferencing condition provide significantly more information about *shape*, through a combination of speech and gesture. Furthermore, with the exception of *material*, all the physical properties and the two body modification properties have large effect sizes ranging from .58 to .85. This emphasises the important contribution of the visual mode for conveying rich semantic information (compare these results to those shown in Appendix D-i where only the verbal mode was considered). These are matched with large *PS* readings, ranging from .85 to 1, underscoring the high degree of nonoverlap between the two population distributions. Our subsequent qualitative analysis seeks to elucidate this distribution. On the other hand, the visual mode makes no contribution for *synonym* and *comparison*, and a small contribution for *judgement* compared to the verbal only comparison (Appendix D-i).

#### *Distribution of referential properties in the videoconferencing condition*

How exactly is information distributed in the verbal and visual modes in the

videoconferencing condition? Is one mode more informative than the other?

The *All properties* distribution shows that learners are significantly more informative in the visual than the verbal mode ( $p = .043$ ), with a large effect size and *PS* (Appendix D-iii). For the physical properties, significantly more semantic information is conveyed visually for *shape* and *size*. The result for *location* also tends towards significance. No differences were statistically established for the other properties. The large effect sizes (from .58 to .64) and very high *PS* readings (.8 to 1) demonstrate the likelihood of the transmission of richer semantic information through the visual mode for *location*, *shape* and *size*. *Process* also has a medium effect size in favour of the visual mode. On the other hand, the verbal mode is likely to be denser in information for *result*, *synonym*, *comparison* and *judgement*.

To summarise the quantitative analyses of the verbal mode in the two conditions, the videoconferencing condition appears richer in referential contributions for *shape*, *process* and *result* (medium to large effect sizes and high *PS* readings). Combining verbal and visual contributions in the videoconferencing condition, the result for *shape* was significant. Although not significant, large effect sizes and high *PS* readings were found for *location*, *position in ear*, *size*, *process* and *result*. Finally, regarding the videoconferencing condition only, the richness of the visual mode was revealed with significant results for *shape* and *size* and large effect sizes and high *PS* readings for *location*, *shape* and *size*. *Process* also had a medium effect size. The verbal mode was richer in information for *result*, *synonym*, *comparison* and *judgement*.

We conclude that, during word search, the visual mode affords extremely rich information for all physical properties of *tunnel earring* except *material*, and for both body modification features.

Although it is to be expected that the affordances provided by the webcam enable semantic contributions to be denser in overall information in the videoconferencing condition, it is an intriguing finding that, when comparing only the verbal mode across the two conditions, the videoconferencing condition is semantically richer, with high *PS* readings and medium to high effect sizes reported for *shape*, *process* and *result*. To better understand why the verbal mode is richer in the videoconferencing condition for semantic information relating to certain referential properties, we now conduct a fine-grained multimodal qualitative analysis.

## **Qualitative analysis**

We investigate the *tunnel earring* word search of one learner from the videoconferencing condition to further account for the quantitative findings above. This is then compared and contrasted to one learner's word search from the audioconferencing condition. We divide the videoconferencing episode into three phases for analysis.

### *Phase 1*

*Figure 1.* Transcription and coding of Phase 1.

**Bold** = referential properties in the visual mode not present in the verbal mode

In the first part of phase 1 (lines 1-7, Figure 1), the learner gazes downwards towards the paper containing the photo being described. Not gazing directly at the teacher signals he is holding his turn and is involved in a self-directed activity. He redirects his gaze towards the teacher in line 8 signalling a transition which seems to constitute a direct invitation for her to participate in the search or ratify her understanding of his description. He then maintains his gaze on her (lines 9-10) but she does not intervene verbally: the teacher listens intently, staring at the learner's image, with a neutral expression<sup>2</sup>.

In line 1, the learner provides a synonym for the target word. The word 'earring' conveys two referential properties which would not enable the teacher to distinguish this generic earring from the target *tunnel earring*. His engagement in the search is displayed by an audible delaying device (line 2), allowing him to hold his turn and engage in a thinking process. Between lines 3 and 7, apart from the repetition of the word 'earring', the iconic gestures communicate significant information, conveying several referential properties, to assist the teacher in interpreting the target item. Indeed, the learner shows the position of the earring, as if

---

<sup>2</sup> Gaze maintenance was shown as a characteristic of word search by Goodwin & Goodwin (1986). Through examples in which the recipient was not initially gazing at the speaker, the researchers illustrated that the recipient gazing towards the speaker was not simply "an accidental type of alignment but something that participants systematically work[ed] to achieve" (1986, p. 54). The researchers also argued that gazing away from the recipient during word search, and producing a 'thinking face' was "not only quite stereotypic and recognizable in many different situations (...); but it has, in fact, been found in other cultures" (1986, p. 57).

he were wearing it himself (cf. Bavelas et al., 2008). Figure 1 illustrates that his gestures carry several referential properties simultaneously and are denser in content than his speech. In a series of gestures positioned near his earlobe, he demonstrates the increase from a small to a large hole, mimicking the gradual stretching of the hole for the tunnel earring. No speech accompanies these gestures. Maintaining his iconic gesture, he adds a deictic gesture (line 9), positioned so as to indicate the position of the tunnel earring directly inside the lobe.

The learner's embodied practices (Mori & Hayashi, 2006; Olsher, 2004) in phase 1 provide visual clues for the teacher, with gestures being "'dense' with information" (Gerwing & Allison, 2011, p. 324). They assist the learner in carrying out forward-oriented repair. By the end of phase 1, the learner's words alone provide insufficient information to identify the specific nature of this earring. His gestures, however, complement his speech, carry a heavier informational load and have an important communicative function. Indeed, at this stage, there is little redundancy between visual and verbal modes.

The first phase of the learner's turn ends with him looking silently towards the screen (line 10). Establishing eye-contact with the teacher and ceasing to gesture, he appears to want to mobilise her participation (cf. Footnote 2). However, she breaks eye contact (line 10), changing her gaze direction and no longer looking at the screen. She shows a thinking face while frowning slightly (Figure 2). There is then a two-second silent pause before Phase 2 begins (Figure 3).

*Figure 2. Thinking face*

## *Phase 2*

*Figure 3. Description of Phase 2.*

The learner seems to interpret the teacher's silence and "conversational facial gesture" (Bavelas et al., 2014, p. 111) as cues to pursue his description, in order to help her comprehend more fully the target item. Phase 3 then commences (Figure 4).

## *Phase 3*

*Figure 4. Transcription and coding of Phase 3.*

**BOLD** = referential properties present in phase 3 but absent in phase 1

\* = referential properties only in the verbal mode

The third phase begins with a self-directed question (line 1), followed by place holders and hesitations (line 2) to gain time for self-repair. Between lines 3 and 14, the learner provides a detailed verbal description of what his gestures had previously illustrated in phase 1. Although there are audible hesitation markers (lines 4, 5, 6 and 9), this description conveys rich semantic information, covering all the referential properties not included in the verbal mode in phase 1 (bold items, Figure 4). While this phase also includes gestures displaying certain properties of the earring, they are less explicit and the gesture span (Beattie & Shovelton, 1999) is smaller than in phase 1.

While gestures continue to convey several referential properties simultaneously, certain gestures are redundant with speech (lines 5, 8 and 14). Nevertheless, they may reinforce the verbal communication, “making a description maximally comprehensible” (Melinger & Levelt, 2004, p. 136). However, the verbal mode conveys referential properties absent in the visual mode (asterisked items in Figure 4: *material; result; judgement*). So, while phase 3 of the word search conveys referential properties in the verbal and visual modes that clarify the target item, the verbal mode is richer.

Let us now compare this word search to an episode from the audioconferencing condition (Figure 5).

Figure 5. Audioconferencing word search episode.

Like the learner in the videoconferencing condition, the learner in the audioconferencing condition begins her description in lines 1 and 2 by providing the synonym ‘earrings’ for the target item. This vague description which conveys two referential properties (*location, synonym*) allows the teacher to gain only a generic representation of the target item. The teacher acknowledges the learner’s contribution by ‘uhu’ in line 3, which seems to encourage the learner to pursue her search. In line 4 the learner’s engagement in the search is shown by an audible hesitation marker which enables her to hold her turn as she pauses to think. She pursues the search in line 5 by repeating the word ‘earring’ once again and qualifying it as being ‘a bit strange’ (*judgement*). This is accompanied by laughter that could be considered either as playing a role in softening the learner’s judgement or as “nervous laughter” used by an anxious L2 learner (Gregersen, 2005, p. 391) as a delaying device. However, she does not add any other complementary features of the referent to her description. In line 6, there is another audible delaying device,

perhaps as the learner attempts to find other ways of describing the target item more precisely. However, subjective understanding is not reached as the learner then immediately abandons the search, by providing a mark of self-admonishment ('I don't know').

So, to sum up the audioconferencing episode analysed here, we observe that just three referential properties are mentioned (*location, synonym, judgement*) in the learner's description, confirming our inferential statistical results. This description does not offer the teacher sufficiently precise information to identify the specific qualities of the referent.

### **Discussion and conclusion**

The present research aimed, firstly, to examine how the semantic information about a target lexical item, *tunnel earring*, was communicated by language learners during word search in a videoconferencing condition compared to an audioconferencing condition; secondly, to determine whether differences existed between the two conditions in how close the learners approached the conventional meaning of the target item.

For our first research question, examining primarily only the verbal content of the lexical explanations in the two conditions, our quantitative analysis identified that the overall comparison of the semantic information (*All properties*) between the two conditions was not significant. However, when referential properties were sampled individually, more references to *shape* in the verbal mode were made in the videoconferencing condition than in the audioconferencing condition. Additionally, for the two body modification properties *process* and *result*, results showed that learners' semantic contributions in the verbal mode were denser in overall information in the videoconferencing condition.

Turning to the role of gestures in communicating the different semantic features of the target lexical item, our initial quantitative analysis showed that by combining verbal and visual modes, learners conveyed more semantic information in the videoconferencing than the audioconferencing condition. The videoconferencing condition allowed learners to provide significantly more referential information about the property *shape* through combining speech and gesture. Additionally, it afforded rich information for *location, position in ear, size, process* and *result*. Second language learners' gestures thus communicated effectively for only certain referential

properties. These results corroborate findings of previous work on native-speaker interactions which adopted a semantic feature approach to account for the qualitative features of gestures and speech (Beattie & Shovelton, 1999; Gerwing & Allison, 2011). They also support findings reported by Nicolaev (2012) and Cappellini (2013) into the importance of gestures in word search in videoconferencing conditions.

The follow-up multimodal qualitative analysis enabled us to explore why the verbal mode was richer in the videoconferencing condition for semantic information relating to certain referential properties; and to gain a deeper understanding of the contribution of gestures in communicating different semantic features. The most interesting finding was that information distribution across the verbal and visual modes differed depending on the word search phase. Indeed, in phase 1, the learner in the videoconferencing condition first embodied the salient physical referential properties of the *tunnel earring* in the visual mode (*location, shape, size, process*), using precise and easily interpretable iconic and deictic gestures, while the verbal mode was much less dense in semantic information. Gestures also enabled the learner to flag his on-going word search, a finding in accord with Gullberg's (1998) study on language learners. In contrast, in phase 3, the verbal mode became much richer semantically: the learner included all the referential features initially absent in the verbal mode. The gestures in this phase, although still present, became smaller and less precise. These findings are consistent with those reported by Gerwing and Bavelas (2004) who demonstrated that, in native-speaker interactions, gestures for a particular referent evolve as information moves from new to given, with gestures changing from large and precise to smaller and vaguer. A fairly high degree of redundancy between gestures and speech was also present in this interaction phase, a finding corroborating previous observations by McCafferty (2002). Compared with the learner in the videoconferencing condition, the learner in the audioconferencing condition drew upon only a small number of referential properties, resulting in only a vague description of the target item.

These qualitative results enabled us to answer our second research question: having access to the interlocutor's image did indeed seem to enable the learner in the videoconferencing condition to move closer, than the learner in the audioconferencing condition, to the conventional meaning of the target item. This, thus, gave the teacher a more precise understanding of the target lexical item.

Our analysis has shown that phases 1 and 3 of the word search differed in terms of information distribution across the verbal and visual modes. How might we account for this?

Gestures in phase 1 may have served to reduce the learner's cognitive load (McCafferty, 2002), allowing him to reflect on the idea he wished to communicate, but not necessarily how to express it verbally. His verbal description was vague, covering only three referential properties (*location, synonym, position in ear*). By contrast, the visual mode included these features, but added information on complementary features – *shape, size and process* – which were repeatedly illustrated. Rarely did the information transmitted in verbal and visual modes overlap. This goes counter to McCafferty's (2002) findings reported earlier which showed that iconic gestures contributed no additional meaning to speech in the L2 learner he studied. So, while these gestures appeared to have an internal cognitive function, allowing the learner to plan his speech, they also had a potentially communicative function.

Having embodied these salient referential properties in phase 1, the learner proceeded, in phase 3, to verbalise what he had previously demonstrated through gestures. Although his gestures continued, they were less explicit and his words conveyed denser semantic information essential for meaning-making. We hypothesise that, having first embodied the salient physical referential properties of the tunnel earring in the visual mode in phase 1, the learner was able to transfer the referential information in phase 3 to produce a rich verbal description of the target item. This finding provides evidence to support McCafferty's (2002) hypothesis that gestures helped the learner in his study to plan his speech during word search, consequently facilitating L2 production.

We do not know whether learners in the audioconferencing condition were gesturing nor, if they were, the form of their gestures. However, even if they were gesturing, with gestures serving an internal cognitive function, since the learners knew that they were not visible to the teacher, the gestures had no communicative function. Did the absence of the visual mode in the audioconferencing condition curtail the amount of detail provided, and notably the number of salient properties to which the learner alluded, in her speech? Compare this to the videoconferencing learner's strategy that appeared to facilitate communication: he 'rehearsed' what he would say subsequently, using gestures framed in the webcam (phase 1), before verbalising this (phase 3).

If this were the case, it would seem that, in the current study, the linguistic affordances (Blin, 2016) provided by the webcam make videoconferencing a better pedagogical option than audioconferencing for online language learning. Indeed, the pushed output produced by the learner in the videoconferencing condition was lexically much richer than that of the learner in the audioconferencing condition that we examined in the qualitative analysis (or, indeed, of any of the four learners in this condition in the quantitative analysis), and this should therefore have been more favourable to language learning. Furthermore, these affordances also seemed to have enabled the learner in the videoconferencing condition to move closer to the conventional meaning of the target item than the learner in the audioconferencing condition.

Referring back to our first set of inferential statistics (Appendix D-i), we observed that the verbal mode was richer in the videoconferencing condition for the referential properties of *shape*, *process* and *result*. Perhaps visually ‘rehearsing’ helped the learner include these properties in his subsequent semantically rich verbal description. The learner in the audioconferencing condition did not draw upon these properties.

Furthermore, the repetition of various referential properties in the videoconferencing condition in the verbal and visual modes was not only beneficial to the learner regarding pushed output, but potentially useful for the interlocutor’s comprehension. In this pedagogical context, the combination of modes might have enabled the teacher to attain a more precise understanding of the target lexical item and have encouraged her to provide it, if known. Additionally, the visual mode seemed to have allowed the teacher to show her degree of understanding, without contributing verbally (phase 2). Her averted gaze, thinking face and slight smile, coupled with the fact that she was holding her earlobe, may have incited the learner to pursue his search. Her facial gestures as “collateral communication” (Bavelas et al., 2014, p. 123) may have demonstrated to him that he was progressing and should elaborate on his description. In contrast, in the audioconferencing condition, the teacher’s ‘uhu’ in the verbal mode might have been interpreted as being more non-committal and, therefore, less encouraging. We hypothesise that the lack of mutual visibility prevented the learner from providing a verbal description which approached the conventional meaning of the target item. Indeed, had the teacher and learner been able to see one another, as was the case in the videoconferencing episode,

they might have been able to gradually co-construct a more precise description of the tunnel earring through the use of co-verbal resources.

Had the interlocutor in the videoconferencing condition been a fellow language learner, exposure to multimodal output should provide rich input for the interactant, potentially promoting language learning and encouraging further interaction. As Brouwer (2003) has noted, word search episodes can offer language learning opportunities when “(a) the other participant is invited to participate in the search, and (b) the interactants demonstrate an orientation to language expertise, with one participant being a novice and the other being an expert” (2003, p. 542).

The current results go counter to findings in Yanguas’s (2010) and Yamada and Akahori’s (2009) studies in which more pushed output was produced in the audioconferencing condition. Differences could be explained by the fact that these studies had learner-learner interactions whereas the current study has focused on teacher-learner interactions. The reassuring and encouraging presence of a teacher was perhaps less face-threatening for the learner who was prepared to elaborate further on his description.

### ***Shortcomings and implications for future research***

The study presents several shortcomings. One limitation is the teacher’s confederate status. In authentic online teaching situations, the teacher as a naïve interlocutor does not necessarily know the learner’s communicative intentions, so interactions may follow a different trajectory. The research design also restricted the teacher’s verbal contributions. Whilst a referential communication task could lead to more balanced interaction, if the same teacher were used throughout, s/he would become familiar with the material, and gradually become more like an experimental confederate. Using different teachers would make comparison of results more difficult. Whilst an ecological approach enables participants to interact freely, without instructional constraints, experimental control would be quickly lost and the resulting variability would again impede comparisons (Bavelas & Chovil, 2006). A further shortcoming concerns our small sample size. Increasing this brings additional challenges to conducting fine-grained qualitative analyses within a quantitative paradigm, including the time required to transcribe and code data.

The current study has compared a word search episode around the lexical item *tunnel earring*. This makes the findings less generalisable as the lexical item has

several specificities: (1) it is a concrete noun; (2) gestures help to describe a number of its key features and (3) the earlobe can be viewed fully via the webcam. Other lexical items would not have such an advantageous profile. So, while the current exploratory study has shown the affordances of the webcam for meaning-making during this particular word search episode, further studies, on other lexical items, must examine whether the contribution of the webcam is replicated.

The authors are currently studying two other lexical elements from the same corpus, one which is more abstract (*funeral*) and another concrete noun but whose embodiment is only partially framed in the webcam (*wheelchair*). We are exploring whether word search follows the same phases as those identified for *tunnel earring*, in which gestures seem to help learners plan their speech and subsequently provide richer verbal semantic information.

It might also be useful in future studies to conduct retrospective self-confrontation interviews, to assess participants' perceptions of their use of the various verbal and visual resources during the interactions. However, since co-verbal resources tend to be used unconsciously during interaction (Tellier, 2012), and may only be unintentional by-products of participants' cognitive processes (Krauss, Dushay, Chen, & Rauscher, 1995), it might be difficult for participants to explain and justify their use reliably.

We would certainly encourage further work that compares L2 interactions using videoconferencing and audioconferencing, to explore the generalisability of our findings, using different samples, settings and tasks. The current study explored the affordances of the webcam in relation to word search. Clearly, it is desirable to examine other analysis units to assess to what extent the webcam may facilitate online L2 interactions.

## References

- Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, 44, 169–188.
- Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, 15, 415–419.
- Bavelas, J. B., & Chovil, N. (2006). Hand gestures and facial displays as part of language use in face-to-face dialogue. In V. Manusov & M. Patterson (Eds.), *Handbook of nonverbal communication* (pp. 97–115). Thousand Oaks, CA: Sage.
- Bavelas, J. B., Gerwing, J., & Healing, S. (2014). Hand and facial gestures in conversational interaction. In T.M. Holtgraves (Ed.), *The Oxford Handbook of Language and Social Psychology* (pp. 111-130). New York: Oxford University Press.
- Bavelas, J. B., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, 58, 495–520.
- Beattie, G., & Shovelton, H. (1999). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica*, 123, 1–30.
- Blin, F. (2016). The theory of affordances. In C. Caws & M.-J. Hamel (Eds.), *Language-Learner Computer Interactions: theory, methodology and CALL applications* (pp. 41-64). Amsterdam: John Benjamins Publishing Company.
- Brouwer, C.E. (2003). Word Searches in NNS-NS Interaction: Opportunities for Language Learning? *The Modern Language Journal*, 87(4), 534–545.
- Cappellini, M. (2013). *Modélisation systémique des étayages dans un environnement de tandem par visioconférence pour le français et le chinois langues étrangères*. Unpublished PhD thesis. Université Lille 3.
- Carroll, D. (2005). Vowel-marking as an interactional resource in Japanese novice ESL conversation. In K. Richards & P. Seedhouse (Eds.), *Applying Conversation Analysis* (pp. 214–234). Hampshire: Palgrave Macmillan.
- Ciekanski, M., & Chanier, T. (2008). Developing online multimodal verbal communication to enhance the writing process in an audio-graphic conferencing environment. *ReCALL*, 20(2), 162-182.
- Cohen, A. A., & Harrison, R. P. (1973). Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Personality and Social Psychology*, 28(2), 276–279.
- Develotte, C., Guichon, N., & Kern, R. (2008). Allo Berkeley? Ici Lyon . . . Vous nous voyez bien?" Etude d'un dispositif de formation en ligne synchrone francoaméricain a` travers les discours de ses usagers. *Alsic (Apprentissage des Langues et Systèmes d'Information et de Communication)*, 11, 129– 156.
- Diaz, R. M., & Berk, L. E. (1992). *Private speech: From social interaction to self-regulation*. New York: Lawrence Erlbaum Associates.
- Egbert, M., Niebecker L., & Rezzara, S. (2004). Inside first and second language speakers' trouble in understanding. In R. Gardner, & J. Wagner (Eds.), *Second language conversations* (pp. 178–200). London: Continuum.

- Emmorey, K., & Casey, S. (2001). Gesture, thought and spatial language? *Gesture*, 1, 35–50.
- Faraco, M., & Kida, T. (1998). Multimodalité de l'interlangue: Geste et interlangue. In S. Santi, I. Guaitella, C. Cavé & G. Konopczynski (Eds.), *Oralité et gestualité* (pp. 635–639). Paris: L'Harmattan.
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics (4<sup>th</sup> edition)*. London: Sage.
- Fritz, C.O., Morris, P.E., & Richler, J.J. (2011). Effect size estimates: current use, calculations and interpretation. *Journal of Experimental Psychology* 141(1), 2–18.
- Gerwing, J., & Allison, M. (2011). The flexible semantic integration of gestures and words: comparing face-to-face and telephone dialogues. *Gesture*, 11(3), 308–329.
- Gerwing, J., & Bavelas, J. (2004). Linguistic influences on gesture's form. *Gesture*, 4(2), 157-195.
- Gregersen, T. (2005). Nonverbal cues: Clues to the Detection of Foreign Language Anxiety. *Foreign Language Annals*, 38(3), 288-400.
- Gregersen, T., Olivares-Cuhat, G., & Storm, J. (2009). An examination of L1 and L2 Gesture Use: What Role does Proficiency Play? *The Modern Language Journal*, 93(2), 195-208.
- Goodwin, M.H., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, 62(1/2), 51–75.
- Guichon, N. & Cohen, C. (2014). The impact of the webcam on an online L2 interaction. *The Canadian Modern Language Review*, 70(3), 331-354.
- Guichon, N. & Wigham, C.R. (2016). A semiotic perspective on webconferencing supported language teaching. *ReCALL*, 28(1), 62-82.
- Gullberg, M. (1998). *Gesture as a communication strategy in second language discourse: a study of learners of French and Swedish*. Lund: Lund University Press.
- Gullberg, M. (2006). Handling discourse: gestures, reference tracking and communication strategies in early L2. *Language Learning*, 56, 155–196.
- Gullberg, M. (2009). Gestures and the development of semantic representations in first and second language acquisition. *Acquisition et Interaction en Langue Etrangère*. Lia 1, 117-139.
- Hammarberg, B. (1998). The learner's word acquisition attempts in conversation. In D. Albrechtsen, B. Henriksen, I. M. Mees, & E. Poulsen (Eds.), *Perspectives on foreign and second language pedagogy* (177–190). Odense, Denmark: Odense University Press.
- Hampel, R., & Hauck, M. (2004). Towards an Effective Use of Audio Conferencing in Distance Language Courses. *Language Learning and Technology*, 8(1), 66-82.
- Hauck M., & Youngs B. (2008). Telecollaboration in multimodal environments: the impact of task design and learner interaction. *Computer Assisted Language Learning*, 21(2), 87-124.
- Hauser, E. (2003). *Corrective recasts' and other-correction of language form in interaction among native and nonnative speakers of English*. Unpublished Ph.D. dissertation, University of Hawai'i.
- Holt, B., & Tellier, M. (2017). Conduire des explications lexicales. In N. Guichon & M. Tellier (Eds.), *Enseigner l'oral en ligne* (pp. 59-90). Paris: Didier.
- Hrastinski, S. (2008). Asynchronous and Synchronous E-Learning. *EDUCAUSE Quarterly*, 31(4), 51–55.

- Jung, K. (2004). L2 vocabulary development through conversation: A conversation analysis. *Second Language Studies*, 23(1), 27–66.
- Kasper, G., & Kellerman, E. (1997). Introduction: approaches to communication strategies. In G. Kasper, & E. Kellerman (Eds.), *Communication strategies: Psycholinguistic perspective* (pp. 1–13). New York: Longman.
- Kern, R. (2014). Technology as Pharmakon: the promise and perils of the internet for foreign language education. *The Modern Language Journal*, 98(1), 340-357.
- Kozar, O. (2016). Perceptions of webcam use by experienced online teachers and learners: a seeming disconnect between research and practice. *Computer Assisted Language Learning*, 29(4), 779-789.
- Krauss, R.M., Dushay, R.A., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology*, 31(6), 533–552.
- Kurhila, S. (2006). *Second Language Interaction*. Amsterdam: John Benjamins Publishing Company.
- Leech, N.L., & Onwuegbuzie, A.J. (2002). A call for greater use of nonparametric statistics. Paper presented at the annual-meeting of the Mid-South Educational Research Association, Chattanooga, TN, November 7, 2002.
- Levy, M., & Stockwell, G. (2006). *CALL dimensions. Options and issues in Computer- Assisted Language learning*. New Jersey: Lawrence Erlbaum Associates.
- McCafferty, S.G. (2002). Gesture and creating zones of proximal development for second language learning. *Modern Language Journal*, 86(2), 192–203.
- McCafferty, S.G. (2004). Space for cognition: gesture and second language learning. *International Journal of Applied Linguistics*, 14(1), 148–165.
- Melinger, A., & Levelt, W. J. M. (2004). Gesture and the communicative intention of the speaker. *Gesture*, 4, 119–141.
- Mori, J., & Hayashi, M. (2006). The achievement of intersubjectivity through embodied completions: a study of interactions between first and second language speakers. *Applied Linguistics* 27, 195–219.
- Neguera, E., & Lantolf, J.P. (2008). The dialectics of gesture in the construction of meaning in second language oral narratives. In S.G. McCafferty and G. Stam (Eds.), *Gesture: second language acquisition and classroom research* (pp. 88-106). New York: Routledge.
- Nicolaev, V. (2012). *L'apprentissage du FLE dans un dispositif vidéographique synchrone : études séquences métalinguistiques*. Unpublished PhD dissertation. Ecole normale supérieure de Lyon. <https://tel.archives-ouvertes.fr/tel-00793185>.
- Norris, S. (2004). *Analyzing multimodal interaction. A methodological framework*. London: Routledge.
- O'Dowd, R. (2006). The use of videoconferencing and e-mail as mediators of intercultural student ethnography. In J. A. Belz & S. L. Thorne (Eds.), *Computer-mediated intercultural foreign language education* (pp. 86–120). Boston: Heinle & Heinle.
- Olsher, D. (2004). Talk and gesture: the embodied completion of sequential actions in spoken interaction. In Gardner, R. and Wagner, J. (Eds.), *Second language conversations* (pp. 221-245). London: Continuum.

- Oxford University Press and University of Cambridge Local Examinations Syndicate. (2001). *Quick Placement Test*. Oxford, UK.
- Plonsky, L., & Oswald, F.L. (2014). How big is 'big'? Interpreting effect sizes in L2 research. *Language Learning*, 64(4).
- Rosell-Aguilar, F. (2007). Changing tutor roles in online tutorial support for open distance learning through audio-graphic SCMC. *The JALT CALL Journal*, 3(1-2), 81–94.
- Satar, M. (2013). Multimodal language learner interactions via desktop videoconferencing within a framework of social presence: Gaze. *ReCALL*, 25(1), 122-142
- Sloetjes, H., & Wittenburg, P. (2008). Annotation by category – ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.
- Smith, B. (2004). Computer-mediated negotiated interaction and lexical acquisition. *Studies in Second Language Acquisition*, 26, 365–398.
- Stam, G. (2006a). Thinking for speaking about motion: L1 and L2 speech and gesture, *International Review of Applied Linguistics in Language Teaching*, 44(2), 145-171.
- Stam, G. (2006b). Changes in patterns of thinking with second language acquisition. Unpublished doctoral dissertation, University of Chicago, Chicago, IL.
- Stam, G., & McCafferty, S.G. (2008). Gesture studies and second language acquisition. In S.G. McCafferty and G. Stam (Eds.). *Gesture: second language acquisition and classroom research* (pp. 3-24). New York: Routledge.
- Tellier, M. (2012). Former à l'étude de la gestuelle : réflexions didactiques. In R. Vion, A. Giacomi and C. Vargas (Eds.). *La corporalité du langage: Multimodalité, discours et écriture. Hommage à Claire Maury-Rouan* (pp.73-85). Aix en Provence: Presses Universitaires de Provence.
- Vetter, A., & Chanier, T. (2006). Supporting oral production for professional purposes in synchronous communication with heterogeneous learners. *ReCALL*, 18(1), 5-23.
- Wigham, C.R. (2017). A multimodal analysis of lexical explanation sequences in webconferencing supported language teaching. In B. O'Rourke & U. Stickler (Eds.) *Special issue of Language Learning in Higher Education: Synchronous communication technologies in language and intercultural learning and teaching in higher education*, 7(1), 81-108.
- Yamada, M., & Akahori, K. (2007). Social Presence in Synchronous CMC-based Language Learning: How does it affect the productive performance and consciousness of learning objectives? *Computer Assisted Language Learning*, 20(1), 37-65.
- 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>

### **Software**

Camtasia: <http://camtasia-for-mac.en.softonic.com>

ELAN: <http://tla.mpi.nl/tools/tla-tools/elan/>

Skype: <http://www.skype.com>

SPSS: <http://www-01.ibm.com/software/fr/analytics/spss/>

## APPENDICES

### Appendix A: Teacher script

Q = teacher

S = student (hypothetical)

Q:  
Hello!

S:  
Hello.

Q:  
Hi, what's your name?

S:  
My name is George.

Q:  
Hi George, how are you?

S:  
I'm fine thank you, (how are you?)

Q:  
(I'm good thanks) Hi I'm Cindy. I'm going to ask you some questions about a set of pictures that you have been handed, I have never seen them, okay?

S:  
Okay....

Q:  
Are you ready?

S:  
... (sigh) yeah...

Q:  
So first could you tell me what is on the first picture?

S:  
Okay,... there is a man, he is talking to two other people, they are at a concert.

Q:  
How can you tell?

S:  
There is a speaker, it's outside, it may be a festival...

Q:  
And what do they seem to be talking about?

S:  
It seems serious

Q:  
Why can you say that?

S:  
They look serious

Q:  
Why, what do their faces look like?

S:  
...

Q:  
....  
Okay, what about the second picture?

S:  
There is an old lady in a hospital

Q:  
How do you know she is in the hospital?

S:  
...

Q:  
And tell me about the third picture?

S:  
It's a funeral

Q:  
What can you see?

S:  
People carrying a coffin

Q:  
Are they burying it?

S:  
...

Q:  
What about the fourth picture?

S:  
It's the picture of a little girl, she is sad.

Q:  
How can you tell? What is her face like?

S:  
She is looking away from the camera...

Q:  
What are the main colours?

S:  
Black and white

Q:  
How do the pictures make you feel?

S:  
...

Q:  
Alright then George, thank you so much for doing this with me!  
I hope it wasn't too boring!

S:  
No, you are great Cindy!

Q:  
Okay then thanks again, have a nice day! What's your next lesson?

S:  
English

Q:  
Enjoy then, goodbye!

S:  
Goodbye!

### **Appendix B: Transcription conventions**

- (.) One second pause. Each . represents one second.
- () Description of an action in the verbal mode e.g., ((coughs)).
- / Rising intonation
- : Sound is extended

XXX The transcriber was not able to decipher the audio.

## Appendix C: Description of referential properties

Referential property type	Referential property	Description
Physical features	Location	The tunnel earring is located on the person's ear.
	Position	The tunnel earring is positioned in the earlobe rather than, for example, in the cartilage at the top of the ear.
	Shape	The tunnel earring is ring-shaped; a cylindrical, hollow piece of jewellery.
	Size	The tunnel earring is unusually big compared to a standard ear piercing.
	Material	The learner suggests that the material from which the tunnel earring is made is wood.
Relating to body modification	Process	A tunnel earring involves stretching the earlobe over time and by gradually upgrading the size of the jewellery, coaxing the hole to fit bigger and bigger tube earrings.
	Result	The large hole left after a tunnel earring is removed is often a permanent body modification.
Other learner strategies	Synonym	The learner uses words having the same or nearly the same meaning. The coders accepted, for example, ear jewel, piercing, earring.
	Comparison	The learners expressed that the item was similar to or different from another item.
	Judgement	The learner expresses a personal reaction to the target item, for example in the verbal mode through phrases such as 'it's pretty impressive' or, in the visual mode, through facial expressions.

## Appendix D: Quantitative results

i. Mean ranks for distribution of properties verbal mode only, Mann-Whitney results, Effect sizes and Probability of superiority (PS)

	Variable	Mean rank Webcam (N=5)	Mean rank Audio (N=4)	Mann-Whitney results	Effect size <i>r</i>	PS[10]
	All properties	5.3	4.62	U = 8.5, z = -.37, p = .730	.12	.58
Physical	Location	4.8	5.25	U = 11, z = .26, p = 1	.09	.45
	Position	5.6	4.25	U = 7, z = -.78, p = .556	.26	.65
	Shape	6.2	3.5	U = 4, z = -1.79, p = .19	.6	.80
	Size	5	5	U = 10, z = 0., p = 1, 0	0	.50
	Material	5.4	4.5	U = 8, z = -.89, p = .73	.3	.60
	Body Modification	Process	5.8	4	U = 6, z = -1.34, p = .413	.45
Result		6.2	3.5	U = 4, z = -1.79, p = .190	.6	.80
Strategies	Synonym	5.1	4.88	U = 9.5, z = -.13, p = .905	.04	.53
	Comparison	5.3	4.62	U = 8.5, z = -.39, p = .73	.13	.58
	Judgement	5	5	U = 10, z = 0, p = 1	0	.50

ii. Mean ranks for distribution of referential properties webcam (verbal + visual) and audio (verbal), Mann-Whitney results, Effect sizes and Probability of superiority (PS)

	<b>Variable</b>	<b>Mean rank Webcam (N=5)</b>	<b>Mean rank Audio (N=4)</b>	<b>Mann-Whitney results</b>	<b>Effect size <i>r</i></b>	<b>PS</b>
	All properties	6.3	3.38	U = 3.5, z = -1.61, p = .11	.49	.54
Physical	Location	6.5	3.12	U = 2.5, z = -1.84, p = .063	.62	.88
	Position	6.4	3.25	U = 3, z = -1.75, p = .111	.58	.85
	Shape	7	2.5	U = 0, z = -2.56, p = .016*	.85	1
	Size	6.6	3	U = 2, z = -1.98, p = .063	.66	.90
	Material	5.4	4.5	U = 8, z = -.89, p = .73	.3	.60
	Body Modification	Process	6.2	3.5	U = 4, z = -1.75, p = .19	.58
Result		6.2	3.5	U = 4, z = -1.79, p = .19	.6	.80
Strategies	Synonym	5.1	4.88	U = 9.5, z = -.13, p = .905	.04	.53
	Comparison	5.3	4.62	U = 8.5, z = -.39, p = .73	.13	.58
	Judgement	5.1	4.88	U = 9.5, z = -.13, p = .905	.04	.53

iii. Medians for distribution of referential properties - webcam verbal and visual, Wilcoxon signed-rank results, Effect sizes and Probability of superiority (PS)

	<b>Variable</b>	<b>Medians Webcam (verbal) (N=5)</b>	<b>Medians Webcam (visual) (N=5)</b>	<b>Wilcoxon signed-rank results</b>	<b>Effect size <i>r</i></b>	<b>PS</b>
	All properties	11	22	T = 15, p = .043*	.64	1
Physical	Location	2	9	T = .000, p = .068	.58	.80
	Position	1	5	T = 6, p = .109	.51	.60
	Shape	2	5	T = 15, p = .043*	.64	1
	Size	1	8	T = 15, p = .042*	.64	1
	Material	0	0	T = .000, p = .317	.32	.20
	Body Modification	Process	0	2	T = 6, p = .102	.52
Result		1	0	T = .000, p = .083	.55	.60
Strategies	Synonym	1	0	T = .000, p = .063	.59	.80
	Comparison	1	0	T = .000, p = .109	.51	.60
	Judgement	1	0	T = 1.5, p = .414	.26	.40

# Tables and Figures

Table 1. Data extent

<b>Condition</b>	<b>No reference to target item (tunnel earring)</b>	<b>Only 'piercing' used</b>	<b>Target item used</b>	<b>Word search episodes around target item</b>
Videoconferencing (N=20)	6	7	2	5
Audioconferencing (N=20)	10	5	1	4

Line	Webcam images	Verbal	Verbal referential properties	Visual referential properties
1		a earring	Location Synonym	
2		you know I donno if you see what I mean		
3		a earring	Location Synonym	Location
4		that's		Location <b>Shape</b> <b>Size</b> <b>Process</b>
5				Location <b>Shape</b> <b>Size</b> Position
6				Location <b>Shape</b> <b>Size</b> <b>Process</b>
7				Location Position <b>Shape</b> <b>Size</b>
8		inside the	Position	Location <b>Shape</b> <b>Size</b>
9		ear actually	Location	Location Position
10				

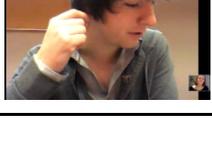
Figure 1. Transcription and coding of Phase 1.



Figure 2. Thinking face

1			<p>The learner maintains his gaze on the teacher, adding nothing in the verbal mode. Her response is purely visual; she maintains her thinking face, looks away from the screen, and moves her right hand to hold her earlobe. This gesture seems to confirm her understanding of the location and position of the target lexical item.</p>
2			<p>He maintains his gaze towards the screen. The teacher holds her gesture and withdrawn gaze, and produces a slight smile.</p>

Figure 3. Description of Phase 2.

	Webcam images	Verbal	Verbal referential properties	Visual referential properties
1		how you say that		
2		er: it's er:		
3		just like an earring	<b>Comparison</b> Location Synonym	
4		that's XXX er a wood ring I think	<b>Material*</b> <b>Shape</b>	
5		er: they start with a small ring	<b>Process</b> <b>Size</b> <b>Shape</b>	Location Shape Size
6		they er:		Location Shape Size
7		place it		Location Position Shape Size Process
8		inside the ear	Position in ear Location	Location Position Shape Size Process
9		and then er:	<b>Process</b>	

10		em		
11		step by step	<b>Process</b>	
12		they put the size up	<b>Process Size</b>	
13		and finally they've got a	<b>Process Result*</b>	
14		pretty big thing inside the ear	<b>Judgement*</b> <b>Size</b> Position Location	Location Position Size Shape

Figure 4. Transcription and coding of Phase 3.

Line	Participant	Verbal	Verbal referential properties
1	Learner	A ( <i>one of the man have</i> ) er (..) earrings	Location Synonym
2	Learner	earrings	Location Synonym
3	Teacher	uhu	
4	Learner	A so er: (...)	
5	Learner	a bit strange earrings ((laughs))	Judgement Location Synonym
6	Learner	so er: I don't know	
<b>Absent semantic features</b>			Position Shape Size Material Process Result Comparison

Figure 5. Audioconferencing word search episode.