



**HAL**  
open science

# A comparison of gesture production in L1 and L2 during video-mediated task-based teletandem interaction

Benjamin Holt

► **To cite this version:**

Benjamin Holt. A comparison of gesture production in L1 and L2 during video-mediated task-based teletandem interaction. *Multimodality across Epistemologies in Second Language Research*, 1, Routledge, pp.153-170, 2024, 9781003355670. 10.4324/9781003355670-11 . hal-04530459

**HAL Id: hal-04530459**

**<https://cnrs.hal.science/hal-04530459>**

Submitted on 3 Apr 2024

**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.

# Multimodality across Epistemologies in Second Language Research

Edited by  
Amanda Brown and Søren W. Eskildsen

First published 2024

ISBN: 9781032409818 (hbk)

ISBN: 9781032409832 (pbk)

ISBN: 9781003355670 (ebk)

---

## 10 A comparison of gesture production in L1 and L2 during video-mediated task-based teletandem interaction

*Benjamin Holt*

CC BY-NC-SA

DOI: 10.4324/9781003355670-11

The funder of the Open Access version of this chapter is



**Routledge**  
Taylor & Francis Group  
NEW YORK AND LONDON

# 10 A comparison of gesture production in L1 and L2 during video-mediated task-based teletandem interaction

*Benjamin Holt*

## Research focus

Video-mediated teletandem interaction, in which two learners of different L1s interact in both languages, has been a popular pedagogical configuration for desktop videoconferencing for at least two decades. This study aims to compare the gesture production (rate, type, and visibility) of teletandem participants as they engage in bilingual French/English task-based video-mediated interaction. We specifically address differences in the participants' gesture production in their two languages (L1 vs L2).

In the next section of this chapter, we present the important role that gestures play in foreign language teaching and learning, and explain how gestures relate to the specific context of desktop videoconferencing environments. Then, we present our data, research questions, and methodology. Finally, we present our findings and conclusion.

## Background

### *Videoconferencing*

Videoconferencing has been used in foreign language classrooms since the early 2000s, first for group-to-group interactions that enabled students to develop intercultural knowledge and skills (O'Dowd, 2016), then for one-on-one communication that enabled learners to gain conversational practice with L1 speakers of their target language. This has been referred to as the teletandem (TT) model (Little & Brammerts, 1996; Telles, 2015). In another popular configuration, based on the *Français en Première Ligne* (F1L) model (Develotte et al., 2008), future foreign language teachers in training gain pedagogical experience by teaching their L1 to learners of that language. Both configurations share similarities that are attributable to the nature of task-based video-mediated communication.

Videoconferencing environments are semiotically rich, allowing researchers to focus on fine-grained aspects of online foreign language teaching from a multimodal perspective. Units of analysis can include

online presence, lexical explanations, feedback, and the management of technical difficulties (Guichon & Tellier, 2017), among others. Gestures that interlocutors make with their hands are sometimes visible thanks to the webcam. Recent studies, adopting the methodological framework of multimodal conversation analysis (Balaman & Pekarek Doehler, 2022; Çolak & Balaman, 2022; Jakonen et al., 2022; Ro, 2023; Uskokovic & Talehgani-Nikazm, 2022), have explored the tight integration of speech with embodied action such as gestures and the use of multimodal onscreen resources as participants navigate the dual and sometimes conflicting priorities of progressivity of talk and progressivity of task.

### *Gesture, multimodality, and video-mediated interaction*

Gesture and speech are part of humans' semiotic resources and must be analyzed together (Kendon, 2004; McNeill, 1992, 2005). Multimodal conversation analysis has shown gestures to contribute to talk and task progressivity in multiple ways, and to be an integral part of bodily action (for a recent review, see Piirainen-Marsh et al., 2022). Gestures are multifunctional; in addition to providing information that is referential (Kendon, 2004) or representational (McNeill, 1992), gestures fulfill important interactional roles, both in face-to-face (Kendon, 2004, 2017) and in video-mediated interaction (Uskokovic & Talehgani-Nikazm, 2022). Although gestures can help speakers find their words and organize their thoughts (Alibali et al., 2000), they also serve communicative purposes and are "meant to be seen" (Alibali et al., 2001, p. 169) as interlocutors tend to produce more gestures when they know that they are visible to an interlocutor. This is true even for video-mediated interaction, where speakers have been shown to gesture more when they know that the webcam is turned on (Mol et al., 2011).

Gestures are of particular importance in foreign language teaching and learning, as evidenced by the rich body of conversation analytic research on classroom interaction (Kunitz et al., 2021; Sert, 2015; Walsh, 2006; Walsh et al., 2011). Gestures have been shown to fulfill important pedagogical functions including transmitting linguistic information, managing the classroom, and providing feedback to learners (Sime, 2008; Tellier, 2008a), and should therefore be included in teacher training programs (Sert, 2015; Tellier & Cadet, 2014; Yerian & Tellier, this volume). Gestures aid comprehension in L2 listening (see Stam & Tellier, 2022 for a review) and reinforce vocabulary learning (Macedonia, 2014; Macedonia & Klimesch, 2014; Macedonia et al., 2019; Macedonia & von Kriegstein, 2012; Quinn-Allen, 1995; Tellier, 2008b), especially when the gestures are reproduced by learners. When explaining words, future foreign language teachers have been shown to use their gestures to transmit rich information pertaining to all aspects of word meaning, both in person (Tellier et

al., 2021) and in video-mediated environments (Holt, 2020, 2021; Holt & Tellier, 2017).

Despite the important role that gestures play in foreign language teaching and learning and in L2 interaction, online teachers do not always make their gestures visible to the webcam. Guichon and Wigham (2016) studied the gesture production of French tutors by comparing a webcam view with an external camera view and found that most of the gestures produced by the tutors were invisible to their distance learners because they were produced outside the webcam's field of view. Given the importance of gestures during communication breakdowns, Holt et al. (2015) hypothesized that these same tutors would increase their gesture rate when there were visible signs of incomprehension. This turned out not to be the case. It therefore seems necessary, as suggested by Guichon and Wigham (2016, p. 67), that online foreign language teachers develop a "critical semiotic awareness" in order to use the tools at their disposal in the most pedagogically effective way possible.

As TT interaction is bilingual by definition, this study compares gesture production across languages. There is some disagreement as to whether bilinguals produce more gestures in their dominant language or in their weaker language (Nicoladis, 2007; Nicoladis & Smithson, 2022). However, L2 speakers have been shown to produce more gestures overall, and more referential gestures, than L1 speakers during disfluent speech (Graziano & Gullberg, 2018). There is some debate as to whether these gestures help the speaker find the word by virtue of the Lexical Retrieval Hypothesis (Krauss et al., 2000) or by lightening the cognitive load (see Nicoladis & Smithson, 2022 for a review), or whether these gestures are meant to elicit help from an L1 listener in finding the word (Gullberg, 2011). To our knowledge, no study has compared gesture production across languages in a video-mediated environment. In light of (1) the unsettled question of whether bilinguals gesture more in their L1 or in their L2, and (2) the usefulness of gesture as a pedagogical resource, this study aims to answer the following questions:

1. Do TT participants produce more gestures in their L1 or in their L2?
2. Do TT participants make more of their gestures visible in their L1 or in their L2?
3. How might this visibility differ?

## **Method**

### *Data source*

Our data are part of the larger VAPVISIO corpus (Cappellini et al., 2023) which aims to directly compare TT interaction and F1L interaction in order to determine which multimodal teaching skills emerge naturally from the

interaction (TT) and which ones require formal training (F1L). For each configuration, two international collaboration projects were implemented between Aix-Marseille University (AMU) and partner universities in the United States, China, and Hong Kong during the spring semester of 2019. Over 70 weekly one-hour sessions were recorded over five weeks. For this study, we have selected two one-hour Skype sessions from the TT project between AMU and Arizona State University (ASU) in order to compare participants' gesture production in French and in English. Both sessions were recorded during the same week, and therefore followed the same activity. In addition to answering discussion questions on environmental sustainability, the participants were asked to analyze and discuss an infographic published by the United Nations promoting sustainable development, as well as an internet meme poking fun at corporate greenwashing. All four participants were female. Our analyses focus on Darie and Océlia, the two French participants.

### *Recording procedure*

The participants completed their weekly Skype sessions individually in order to benefit from a specialized setup in the language center at AMU. The dedicated room featured a laptop computer connected to an external mouse, keyboard, and monitor with a Tobii eye-tracking device. An external sound card, connected to a separate desktop computer, was used for recording the sound. An external camera was used to capture the gestures and other artefacts that were not visible to the webcam. Due to the Tobii eye-tracking device, the participants were required to sit between 50 and 90 centimeters from the screen. They could also see their own webcam image thanks to the “rear-view mirror effect” (Guichon, 2017, p. 35).

### *Data annotation*

To answer our research questions, we first transcribed all of the speech using the SPPAS transcription convention (Bigi, 2012) and the Praat transcription tool. This convention was chosen because SPPAS was used for other studies that are part of the VAPVISIO project. We then annotated all visible gestures using ELAN (Wittenburg et al., 2006). For this we chose a McNeillian annotation scheme (McNeill, 1992, 2005) which enables comparison with other studies (Holt et al., 2015; Tellier et al., 2021). As Stam and Buescher (2018) point out, such cross-study comparison is often difficult in gesture studies. We annotated gestures according to their semantic properties: Iconic (which describe physical characteristics of the referent), metaphoric (which represent abstract concepts such as ideas), deictic (pointing), and beat (rhythmic without semantic content) gestures. We also added emblems (Ekman & Friesen, 1969) which are gestures that have a

specific form and meaning understood by a culture such as the thumbs-up sign, as well as non-identifiable gestures which were unrecognizable due to lack of visibility. We annotated the gestures from each session twice, once from the external camera’s point of view and once from the webcam’s point of view.

After annotating the gestures, we calculated the gesture rate (number of gestures per 1,000 tokens) for all participants, for both languages and from both camera views. Throughout some of the sessions, the French participants’ webcam image was covered up by another window on the screen such as a document or webpage, making it impossible to observe their gestures from the webcam’s point of view. We therefore excluded these sections of the recordings (both the webcam recording and the corresponding external camera recording) when establishing overall word and gesture counts so as not to falsify the results.

In order to answer our third research question concerning gesture visibility, we elaborated an annotation scheme in order to classify the ways in which the gestures differed between the two camera views. Figure 10.1 illustrates our annotation scheme, defined as follows:

- *Completely visible* gestures are those that were annotated the same way twice – once when watching the external camera recording and once when watching the webcam recording – and whose stroke (Kendon, 2004) is completely visible to the webcam. The gesture in the example below was coded as “metaphoric” both times because the hands represent the abstract concept of “the thing for the children” referring to one of UNESCO’s programs.
- *Reduced* gestures are those that were annotated the same way twice, but whose stroke is not completely visible to the webcam. The gesture in the example below was annotated as “iconic” both times because the hands









| Type            | Completely visible                                                                  | Reduced                                                                             | Modified                                                                            | Completely invisible                                                                |
|-----------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| External camera |  |  |  |  |
| Webcam          |  |  |  |  |
| Speech          | “the thing”                                                                         | “a little notebook”                                                                 | “we have windmills”                                                                 | “that’s what people...”                                                             |

Figure 10.1 Annotation of gesture visibility.

represent the physical form of a notebook being opened, but since only the thumbs are visible to the webcam, it was labeled as “reduced”.

- *Modified* gestures are those that were annotated differently depending on the camera view. This can be caused by the webcam’s lack of peripheral visibility, lack of depth perspective, or both. The gesture in the example above was annotated as an “emblem” (counting gesture on the fingers) when watching the external camera recording, but as “non-identifiable” when watching the webcam recording because only the tip of one finger was visible.
- A *completely invisible* gesture is one that was annotated when watching the external camera recording but not when watching the webcam recording. In the example above the gesture was coded as “deictic” from the external camera’s point of view but was not annotated at all when watching the webcam recording (see Figure 10.1).

## Findings

### *Do TT participants produce more gestures in their L1 or in their L2?*

After viewing our recordings multiple times, we annotated 1,227 gestures from the point of view of the external camera. Next, we calculated the gesture rate per 1,000 words for each type of gesture and for each language. The top part of Figure 10.2 shows the gesture rate per person, per type of gesture, and per language.

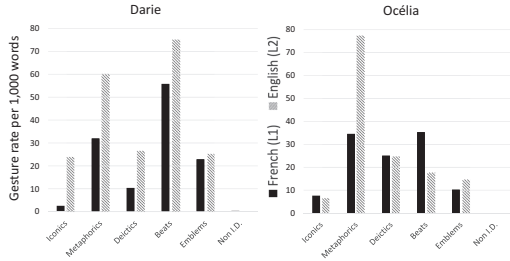
We first notice that both Darie and Océlia gesture more in their L2 (English) than they do in their L1 (French). Darie’s overall rates are 124 gestures per 1,000 words in French and 212 in English. Océlia’s overall rates are 113 for French and 142 for English. This finding corroborates previous research that has pointed to an inverse relationship between linguistic proficiency and gesture rate (Gullberg, 2006; Nicoladis, 2007). Another finding is that both speakers produce more metaphoric gestures in their L2 than in their L1, whereas the production of emblems appears to be similar in both languages for both speakers. There are some interindividual differences as well: Darie produces more iconic, deictic, and beat gestures in her L2 than she does in her L1, whereas for Océlia there is either little difference (iconics and deictics) or the opposite is true (beats).

### *Do TT participants make more of their gestures visible in their L1 or in their L2?*

In order to answer our second research question, we used the screen recordings to annotate 677 gestures from the webcam’s point of view. We then compared these annotations with those from the external camera’s point of view to calculate differences in gesture rates and changes in visibility. For both participants, both languages, and all gesture types, the gesture



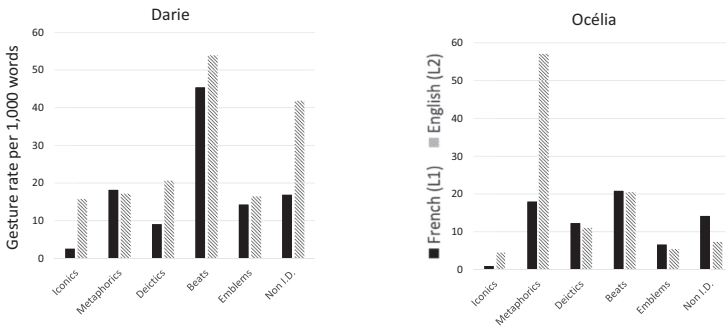
RQ1: Gesture rates and types from the external camera's point of view



RQ2: Overall gesture rate per language and per point of view

| Participant | Language | Gesture rate (external camera) | Gesture rate (webcam) | % Change in visibility |
|-------------|----------|--------------------------------|-----------------------|------------------------|
| Darie       | French   | 124                            | 106                   | -14.5%                 |
|             | English  | 212                            | 166                   | -21.7%                 |
| Océlia      | French   | 113                            | 73                    | -35.4%                 |
|             | English  | 142                            | 107                   | -24.6%                 |

RQ2: Gesture rates and types from the webcam's point of view



RQ3: Gesture modification

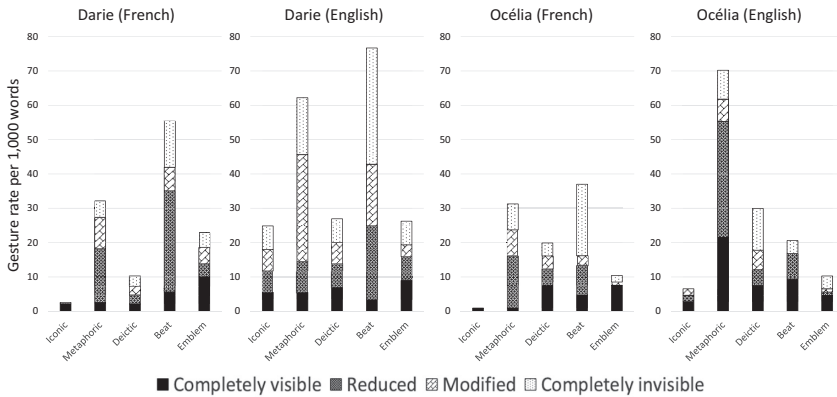


Figure 10.2 Gesture rates, gesture types, and gesture modification.

rates calculated from the webcam's point of view are lower than those calculated from that of the external camera, indicating loss in terms of gesture visibility. The table in Figure 10.2 shows the overall gesture rates for each language and the percentage change when switching from the external camera to the webcam. The percentage change takes into account the number of gestures that are visible but does not speak to the degree to which each gesture is visible. This is addressed below when answering our third research question.

As expected, gesture rates are lower overall when switching from the external camera to the webcam. However, most of the gestures produced by our subjects are still visible, albeit to varying degrees. This contrasts with findings from Guichon and Wigham's (2016) study in which most of the gestures were invisible to the webcam (no exact percentages were given). Overall, it seems that although participants can see their own webcam image, they do not constantly monitor it and often fail to make their gestures visible to their interlocutor. Our data show that Darie makes her gestures slightly more visible in French than in English, whereas Océlia makes her gestures slightly more visible in English than in French.

Next, we calculated the visibility rates by gesture type. The table in Figure 10.2 shows the gesture rates for each participant, language, and type of gesture from the webcam's point of view. A comparison of the first two bar charts in Figure 10.2 reveals that when switching camera views, the overall trends remain the same for Darie's production of iconic, deictic, beat, and emblematic gestures, and for Océlia's production of metaphoric, deictic, and emblematic gestures. Concerning the differences when switching camera views, Darie's metaphoric gestures become much less visible in English, whereas Océlia's iconic and beat gestures become much less visible in French. Finally, both participants produce more non-identifiable gestures from the webcam's point of view, with Darie producing more in English and Océlia producing more in French.

The finding that both participants gesture more in English than they do in French may have to do with attempts to elicit help from the L1 speaker (Gullberg, 2011) and/or to lighten the cognitive load (Nicoladis & Smithson, 2022). Two examples illustrate this.

In the first example, Darie is describing a fundraising event where she ran laps around a track to collect donations to fight hunger. When she encounters lexical difficulties, she opens Google Translate to search for several words including "flyer", "notebook", and "sponsor". When it comes time to translate the French word "tour", she makes verbal and non-lexical vocalizations ("how" and laughing) (Balaman & Pekarek Doehler, 2022) before looking up the word online in order to advance the talk and the task (Çolak & Balaman, 2022). The first result offered by Google Translate is "tower", but the correct word for this context, "lap", is the fifth item down



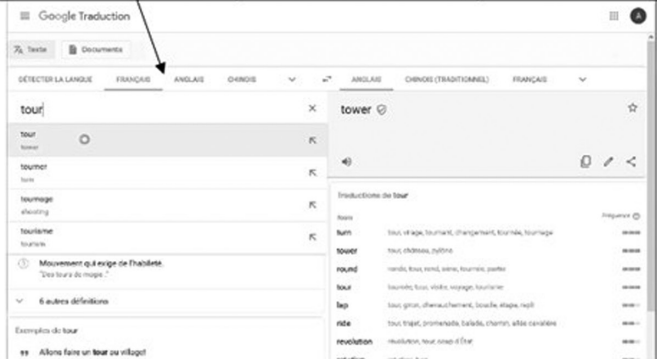


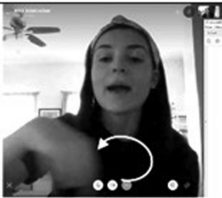
| Speech                                                                                                                                   | Screen or webcam                                                                    | External camera                                                                   | Observations                                                                                                                                        |
|------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| DAR: euh we explain them + euh (0.7) um like (0.3) if we (0.4) if we run (0.6) run (0.3) ran (0.6) one (1.5) tour (0.2) co-@ how @ (0.9) |    |  | Darie looks up the translation of “tour” on Google Translate. The first result is “tower” ( <i>une tour</i> ), instead of “lap” ( <i>un tour</i> ). |
|                                                                                                                                          |    |                                                                                   |                                                                                                                                                     |
| DAR: run (1.1)                                                                                                                           |    |  | Darie makes a twirling motion with her finger while searching for the word “lap.” The gesture is not completely visible to the webcam.              |
| JOS: yeah like one lap<br>DAR: yeah + thank you @ (0.8) one lap um                                                                       |  |                                                                                   | Josephine repeats Darie’s iconic gesture.                                                                                                           |

Figure 10.3 Darie elicits help with an iconic gesture.

the list. Darie produces an iconic gesture for “lap”, which enables Josephine, her interlocutor, to suggest the correct word while reproducing the same iconic gesture (see de Fornel, 1992; Majlesi, 2022 for previous work on return gestures). Even though Darie’s gesture is not completely visible to the webcam, Josephine has no trouble interpreting it and repeating it. The word “lap” is accepted by Darie and the talk continues (see Figure 10.3).

In the second example (Figure 10.4), Océlia and her partner are discussing the reputation of feminism. Océlia gestures at a high rate as she





| Océlia's speech                                                                                                                                                                         | Webcam                                                                            | External camera                                                                   | Observations                         |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|--------------------------------------|
| well that's the k- the kind of girls that <b>will go topless you know</b> in + outside and they will scream and they will do                                                            |  |  | Completely visible iconic gesture.   |
| (0.2) like dangerous thing (0.3) for themselves and for others (1.0) so that's (0.4) very (0.3) the one that + we don't really (0.3) see as our (0.5) <b>um + image</b> + of + feminism |  |  | Completely visible metaphoric image. |

Figure 10.4 Océlia makes her gestures visible.

speaks, possibly to lighten her cognitive load. Due to space, only two of her gestures are represented: An iconic gesture as she says, “go topless”, and a metaphoric gesture as she says “image”. She produces the metaphoric image twice: Once as she says “um” while searching for the word, and once again as she says the word “image”. Similar to Darie, she produces gestures as she searches for her words, but Océlia is able to find the words without the help of her partner. It therefore seems that Darie and Océlia gesture in order to elicit help and to lighten their cognitive loads (Nicoladis & Smithson, 2022). These examples illustrate the need for a more thorough investigation of why and in what interactional contexts the participants gesture. This remains a point for future research.

#### *How might gesture visibility differ?*

In addition to calculating the overall visibility rates as binary data (visible or not visible), we used the scheme presented in the methodology section to annotate visibility changes between camera views for each gesture dimension. The gesture rates and visibility rates per 1,000 words for each speaker, language, and gesture dimension are represented in the final bar chart of Figure 10.2. The vertical bars differ slightly in height from those represented in the first bar chart, because whereas that one takes into account all gestures recorded by the external camera for the entire interaction, this one only looks at the comparable parts of the interaction during which the speaker's webcam image was not covered up. We also do not include non-identifiable gestures since so few gestures were annotated this way from the external camera's point of view.

The first thing we notice is that the vertical bars are mostly evenly divided between “completely visible” and “reduced” on the bottom, and “modified” and “completely invisible” on the top. In other words, our

annotation scheme appears to have split these bars in half, suggesting a linear, gradual progression from completely visible to completely invisible. Next, we notice that some gesture types are more “visible” than others, such as iconics and emblems. The percentage may have been different if more of Océlia’s iconic gestures had been comparable (produced without her webcam image obscured). Consistent with the trend mentioned above, not only does Océlia produce more metaphoric gestures in English than she does in French, but a greater proportion of them are visible in her L2. The visibility of deictics and beats do not seem to change much for either speaker when switching languages, although Océlia’s beats are slightly more visible in English than in French.

## **Conclusion**

This chapter set out to compare the gesture production from two different camera views of two French TT participants who engaged in video-mediated task-based interaction for one hour in French and in English. Our first finding is that depending on speaker and language, 15–35% of gestures are not visible to the webcam. This corroborates previous studies (Guichon & Wigham, 2016) that have found significant loss of gesture visibility from the webcam’s point of view. We therefore suggest that TT participants be made explicitly aware of the effect that the webcam can have on gesture visibility. This need not include full-blown auto- or hetero-analysis of recordings as is sometimes done in teacher training programs (Azaoui, 2022; Gadoni & Tellier, 2014); a future study could involve comparison of two groups of TT participants, one of which is instructed beforehand to pay special attention to the framing of their gestures. This would allow us to discern if overall webcam gesture visibility has more to do with status (future teacher or not), individual communication style, or simply the instructions that are provided in advance.

Next, we found that both participants gesture more in their L2 than in their L1, which corroborates the finding mentioned by various researchers (Gullberg, 2006; Nicoladis, 2007) that bilingual speakers tend to gesture more in their non-dominant language. The two participants that we have studied seem to be using their gestures not to help their interlocutors understand French (Tellier et al., 2021), but to facilitate their output in English. Although both participants produce more gestures overall in their L2 than in their L1, there are some differences between the two participants. Darie produces more iconics, deictics, and beats in her L2 than she does in her L1, and this is true regardless of camera view. She therefore aligns with what some studies would describe as a “typical” profile of a bilingual speaker who has a significant gap in proficiency between their two languages (see Nicoladis, 2007 for a review). As shown in the example above, these iconic gestures could enable Darie, who has a slightly lower

level of English than Océlia, to elicit her partner's help in finding words (Graziano & Gullberg, 2018). Not all of Darie's iconic gestures in English are completely visible, which could mean that she is not fully aware of the effect that the webcam has on gesture visibility, or of the fact that she is using a gesture/speech combination to elicit help from her interlocutor. However, as shown in the example above, this imperfect visibility does not prevent her interlocutor from interpreting and repeating the gesture in order to overcome the communicational difficulty (Gullberg, 2011; Eskildsen & Wagner, 2013). Océlia has a slightly higher level of English and does not rely as much on her partner for help. Her L2 gestures therefore most likely serve to reduce her cognitive load while speaking. These differences could be related to language proficiency, but also to individual communicative style. As Gullberg (2011, p. 145) points out, "individual communicative styles appear to determine behavior at least as much as the difficulties experienced".

Lastly, differences were observed concerning visibility rates in each language. As mentioned, Océlia's gestures are slightly more visible in English, and Darie's are slightly more visible in French. However, this difference is not great enough to warrant a claim that Darie's higher L1 gesture rate comes from a desire to help her American interlocutor understand French, especially since Josephine is very competent in French (Darie has said as much) and Darie makes little effort to engage in foreigner talk (Ferguson, 1975). Furthermore, if we were to rank the gesture types shown in the final bar chart by visibility rates in French, they would not align with the order of importance of gesture types for L2 learning given by Nicoladis (2007): Iconics, beats, deictics, then emblems. Future research should therefore compare gesture visibility rates of TT participants to those of future teachers in training (F1L) in order to see if online teachers naturally or consciously increase the visibility of the types of gestures that are most important for L2 learning.

This study has several limitations that need to be addressed. The first is that we only studied two pairs of interlocutors during one session. More data are needed to draw any generalizable conclusions about gesture visibility and language. We hope that our annotation scheme for gesture visibility will be reused in future studies. Another point is that we counted gesture production for the entire interaction instead of focusing on certain types of sequences such as word searches (Uskokovic & Taleghani-Nikazm, 2022) or instances of corrective feedback (Inceoglu & Loewen, 2022). If we had isolated specific types of sequences, our results may have been different. Another technical limitation is that during some parts of the interaction, the participants' webcam images were covered up on the screen recording, meaning that comparison of gesture visibility was not possible during these parts of the interaction. This is why so few of Océlia's iconic

gestures are comparable in English. It would be useful if Skype and other videoconferencing platforms could record the webcam view independently of what is on the screen. Finally, we chose a McNeillian annotation scheme so that our results could be easily compared to those of other studies in the field. As pointed out by Urbanski and Stam (2022), results are influenced and constrained by coding schemes, and an alternate classification of gestures – for example by pedagogical (Tellier, 2008a) or pragmatic (Kendon, 2017) function – could have yielded different results.

## 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(2), 169–188. <https://doi.org/10.1006/jmla.2000.2752>
- Alibali, M. W., Kita, S., & Young, A. J. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes*, 15(6), 593–613. <https://doi.org/10.1080/016909600750040571>
- Azaoui, B. (2022). Analyzing metagestuality. In G. Stam & K. Urbanski (Eds.), *Gesture and multimodality in second language acquisition: A research guide* (pp. 48–72). Routledge. <https://doi.org/10.4324/9781003100683-4>
- Balaman, U., & Pekarek Doehler, S. (2022). Navigating the complex social ecology of screen-based activity in video-mediated interaction. *Pragmatics*, 32(1), 54–79. <https://doi.org/10.1075/prag.20023.bal>
- Bigi, B. (2012). SPPAS: A tool for the phonetic segmentations of Speech. In N. Calzolari, K. Choukri, T. Declerck, M. Uğur Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the eighth international conference on language resources and evaluation* (pp. 1748–1755). European Language Resources Association (ELRA).
- Cappellini, M., Holt, B., Bigi, B., Tellier, M., & Zielinski, C. (2023). A multimodal corpus to study videoconference interactions for techno-pedagogical competence in second language acquisition and teacher education. *Corpus*. <https://doi.org/10.4000/corpus.7440>
- Çolak, F., & Balaman, U. (2022). The use of online dictionaries in video-mediated L2 interactions for the social accomplishment of virtual exchange tasks. *System*, 106, 102772. <https://doi.org/10.1016/j.system.2022.102772>
- de Fornel, M. (1992). The return gesture: Some remarks on context, inference, and iconic gesture. In P. Auer & A. Di Luzio (Eds.), *Contextualization of language*. John Benjamins. <https://doi.org/10.1075/pbns.22.11for>
- Develotte, C., Guichon, N., & Kern, R. (2008). “Allo Berkeley ? Ici Lyon... Vous nous voyez bien ?” Étude d’un dispositif de formation en ligne synchrone franco-américain à travers les discours de ses usagers. *Alsic*, 11(2), 129–156. <https://doi.org/10.4000/alsic.892>
- Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, and coding. *Semiotica*, 1(1), 49–98. <https://doi.org/10.1515/semi.1969.1.1.49>

- Eskildsen, S. W., & Wagner, J. (2013). Recurring and shared gestures in the L2 classroom: resources for teaching and learning. *European Journal of Applied Linguistics*, 1(1). <https://doi.org/10.1515/eujal-2013-0007>
- Ferguson, C. (1975). Towards a characterization of English foreigner talk. *Anthropological Linguistics*, 17, 1–14.
- Gadoni, A., & Tellier, M. (2014). Conscientisation de la gestuelle pédagogique en formation initiale. In M. Tellier & L. Cadet (Eds.), *Le corps et la voix de l'enseignant: Théorie et pratique* (pp. 207–217). Maison des Langues.
- Graziano, M., & Gullberg, M. (2018). When speech stops, gesture stops: Evidence from developmental and crosslinguistic comparisons. *Frontiers in Psychology*, 9, 879. <https://doi.org/10.3389/fpsyg.2018.00879>
- Guichon, N. (2017). Se construire une présence pédagogique en ligne. In N. Guichon & M. Tellier (Eds.), *Enseigner l'oral en ligne: une approche multimodale* (pp. 29–58). Les Éditions Didier.
- Guichon, N., & Tellier, M. (Eds.). (2017). *Enseigner l'oral en ligne: Une approche multimodale. Les éditions*. Didier. <https://doi.org/10.14375/np.9782278090600>
- Guichon, N., & Wigham, C. R. (2016). A semiotic perspective on webconferencing-supported language teaching. *ReCALL*, 28(1), 62–82. <https://doi.org/10.1017/s0958344015000178>
- Gullberg, M. (2006). Some reasons for studying gesture and second language acquisition (Homage à Adam Kendon). *IRAL - International Review of Applied Linguistics in Language Teaching*, 44(2). <https://doi.org/10.1515/IRAL.2006.004>
- Gullberg, M. (2011). Multilingual multimodality: Communicative difficulties and their solutions in second language use. In J. Streeck, C. Goodwin, & C. LeBaron (Eds.), *Embodied interaction: Language and body in the material world* (pp. 137–151). Cambridge University Press.
- Holt, B. (2020). Le rôle des gestes dans les explications lexicales par visioconférence. *TIPA. Travaux interdisciplinaires sur la parole et le langage*, 36. <https://doi.org/10.4000/tipa.3458>
- Holt, B. (2021). Explaining different aspects of word knowledge: A multimodal analysis of lexical explanation sequences during online French tutoring sessions. *Lexis*. <https://doi.org/10.4000/lexis.5889>
- Holt, B., & Tellier, M. (2017). Conduire des explications lexicales. In N. Guichon & M. Tellier (Eds.), *Enseigner l'oral en ligne* (pp. 59–90). Didier.
- Holt, B., Tellier, M., & Guichon, N. (2015). The use of teaching gestures in an online multimodal environment: The case of incomprehension sequences. In G. Ferré & T. Mark (Eds.), *Gesture and speech in interaction* (4th ed., pp. 149–154). HAL Open Science. <https://hal.science/hal-01195646>
- Inceoglu, S., & Loewen, S. (2022). Analyzing nonverbal corrective feedback. In G. Stam & K. Urbanski (Eds.), *Gesture and multimodality in second language acquisition: A research guide* (pp. 26–47). Routledge. <https://doi.org/10.4324/9781003100683-2>
- Jakonen, T., Dooley, M., & Balaman, U. (2022). Interactional practices in technology-rich L2 environments in and beyond the physical borders of the classroom. *Classroom Discourse*, 13(2), 111–118. <https://doi.org/10.1080/19463014.2022.2063547>



- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511807572>
- Kendon, A. (2017). Pragmatic functions of gestures: Some observations on the history of their study and their nature. *Gesture*, 16(2), 157–175. <https://doi.org/10.1075/gest.16.2.01ken>
- Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). Cambridge University Press. <https://doi.org/10.1017/cbo9780511620850.017>
- Kunitz, S., Markee, N., & Sert, O. (Eds.). (2021). *Classroom-based conversation analytic research: Theoretical and applied perspectives on pedagogy*. Springer. <https://doi.org/10.1007/978-3-030-52193-6>
- Little, D., & Brammerts, H. (Eds.). (1996). *A guide to language learning in tandem via the Internet*. CLCS Occasional paper no. 46. Center for Language and Communication Studies.
- Macedonia, M. (2014). Bringing back the body into the mind: Gestures enhance word learning in foreign language. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01467>
- Macedonia, M., & Klimesch, W. (2014). Long-term effects of gestures on memory for foreign language words trained in the classroom. *Mind, Brain, and Education*, 8(2), 74–88. <https://doi.org/10.1111/mbe.12047>
- Macedonia, M., & von Kriegstein, K. (2012). Gestures enhance foreign language learning. *Biolinguistics*, 6, 393–416. <https://doi.org/10.5964/bioling.8931>
- Majlesi, A. R. (2022). Gestural matching and contingent teaching: Highlighting learnables in table-talk at language cafés. *Social Interaction. Video-Based Studies of Human Sociality*, 5(1). <https://doi.org/10.7146/si.v5i2.130871>
- McNeill, D. (1992). *Hand and mind*. The University of Chicago Press. <https://doi.org/10.7208/chicago/9780226514642.001.0001>
- McNeill, D. (2005). *Gesture & thought*. The University of Chicago Press. <https://doi.org/10.7208/chicago/9780226514642.001.0001>
- Mol, L., Kraemer, E., Maes, A., & Swerts, M. (2011). Seeing and being seen: The effects on gesture production. *Journal of Computer-Mediated Communication*, 17(1), 77–100. <https://doi.org/10.1111/j.1083-6101.2011.01558.x>
- Nicoladis, E. (2007). The effect of bilingualism on the use of manual gestures. *Applied Psycholinguistics*, 28(3), 441–454. <https://doi.org/10.1017/s0142716407070245>
- Nicoladis, E., & Smithson, L. (2022). Gesture in bilingual language acquisition. In A. Morgenstern & S. Goldin-Meadow (Eds.), *Gesture in language: Development across the lifespan* (pp. 297–315). De Gruyter Mouton. <https://doi.org/10.1037/0000269-012>
- O'Dowd, R. (2016). Emerging trends and new directions in telecollaborative learning. *CALICO Journal*, 33(3), 291–310. <https://doi.org/10.1558/cj.v33i3.30747>
- Piirainen-Marsh, A., Lilja, N., & Eskildsen, S. W. (Eds.). (2022). Bodily practices in action formation and ascription in multilingual interaction. *Social Interaction – Video-Based Studies of Human Sociality*, 5(1). <https://tidsskrift.dk/socialinteraction/issue/view/9565>

- Quinn-Allen, L. (1995). The effects of emblematic gestures on the development and access of mental representations of French expressions. *The Modern Language Journal*, 79(4), 521–529. <https://doi.org/10.1111/j.1540-4781.1995.tb05454.x>
- Macedonia, M., Repetto, C., Ischebeck, A., & Mueller, K. (2019). Depth of encoding through observed gestures in foreign language word learning. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.00033>
- Ro, E. (2023). Topicalization as a practice for facilitating L2 discussion in video-mediated extensive reading book club meetings. *System*, 113, 102996. <https://doi.org/10.1016/j.system.2023.102996>
- Sert, O. (2015). *Social interaction and L2 classroom discourse*. Edinburgh University Press. <https://doi.org/10.1515/9780748692651>
- Sime, D. (2008). “Because of her gesture, it’s very easy to understand” – Learners’ perceptions of teachers’ gestures in the foreign language class. In S. McCafferty & G. Stam (Eds.), *Gesture: Second language acquisition and classroom research* (pp. 259–279). Routledge.
- Stam, G., & Buescher, K. (2018). Gesture research. In A. Phakiti, P. De Costa, L. Plonsky & S. Starfield (Eds.), *The Palgrave handbook of applied linguistics research methodology* (pp. 793–809). Palgrave Macmillan. [https://doi.org/10.1057/978-1-137-59900-1\\_36](https://doi.org/10.1057/978-1-137-59900-1_36)
- Stam, G., & Tellier, M. (2022). Gesture helps second and foreign language learning and teaching. In A. Morgenstern & S. Goldin-Meadow (Eds.), *Gesture in language: Development across the lifespan* (pp. 335–363). De Gruyter Mouton. <https://doi.org/10.1037/0000269-014>
- Telles, J. A. (2015). Learning foreign languages in teletandem: Resources and strategies. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, 31(3), 603–632. <https://doi.org/10.1590/0102-4450226475643730772>
- Tellier, M. (2008a). Dire avec des gestes. *Le français dans le monde: Recherche et applications*, 44, 40–50.
- Tellier, M. (2008b). The effect of gestures on second language memorisation by young children. *Gesture*, 8(2), 219–235. <https://doi.org/10.1075/bct.28.06tel>
- Tellier, M., & Cadet, L. (Eds.), (2014). *Le corps et la voix de l’enseignant: Théorie et pratique*. Éditions maison des langues.
- Tellier, M., Stam, G., & Ghio, A. (2021). Handling language: How future language teachers adapt their gestures to their interlocutor. *Gesture*, 20(1), 30–62. <https://doi.org/10.1075/gest.19031.tel>
- Urbanski, K., & Stam, G. (2022). Overview of multimodality and gesture in second language acquisition. In G. Stam & K. Urbanski (Eds.), *Gesture and multimodality in second language acquisition: A research guide* (pp. 1–25). Routledge. <https://doi.org/10.4324/9781003100683-1>
- Uskokovic, B., & Taleghani-Nikazm, C. (2022). Talk and embodied conduct in word searches in video-mediated interactions. *Social Interaction – Video-Based Studies of Human Sociality*, 1(1). <https://doi.org/10.7146/si.v5i2.130876>
- Walsh, S. (2006). *Investigating classroom discourse*. Routledge. <https://doi.org/10.4324/9780203015711>
- Walsh, S., Morton, T., & O’Keeffe, A. (2011). Analysing university spoken interaction: A CL/CA approach. *International Journal of Corpus Linguistics*, 16(3), 325–345. <https://doi.org/10.1075/ijcl.16.3.03wal>

- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In N. Calzolari, Kh. Choukri, A. Gangemi, B. Maegaard, J. Mariani, J. Odijk, & D. Tapias (Eds.), *Proceedings of the 5th international conference on language resources and evaluation (LREC 2006)* (pp. 1556–1559). European Language Resources Association (ELRA).



**Taylor & Francis**

Taylor & Francis Group

<http://taylorandfrancis.com>