Embodied cognition in the booth

Referential and pragmatic gestures in simultaneous interpreting

Celia Martín de León and Alba Fernández Santana
Universidad de Las Palmas de Gran Canaria

From an enactivist perspective, cognition can be described as embodied, since it is determined by our bodily, multisensory, affective interaction with the environment, in particular by our social interaction. In recent years, interpreting has been defined as a multimodal, embodied cognitive activity of inter-lingual mediation, and research on gestures in conference interpreting has found that simultaneous interpreters, although not visible for their audience, do gesture in the booth. However, gestures in interpreting are yet understudied. This paper presents an exploratory, in-depth descriptive study with the aim of generating hypotheses about the cognitive functions of gestures in simultaneous interpreting. To this end, we investigate the different types of gesture that emerge throughout a whole process of simultaneous interpreting, in conjunction with the concurrent speech, the interpreter's interaction with her environment and her own description of her production of mental images and gestures. The research question guiding our investigation is: What functions do the different types of gestures play in the interpreting process? The results suggest that, in the analyzed material, referential gestures tend to support the construction of meanings, while the main role of pragmatic gestures consists in helping to manage the progress of the interpreting process.

Keywords: cognition, enactivism, embodiment, simultaneous interpreting, gestures

1. Introduction

In cognitive science, the idea that cognition is embodied has been gaining ground in recent decades. However, the notion of embodiment has been approached from different perspectives ranging from conservative models (close to classic computational views) to more radical conceptions that emphasize that perception and

https://doi.org/10.1075/cogls.00079.mar
Cognitive Linguistic Studies 8:2 (2021), pp. 277–306. ISSN 2213-8722 | E-ISSN 2213-8730

cognition are oriented to action (Gallagher 2017: 26–47). In this paper, we adopt an enactivist perspective that understands cognition as composed by "dynamical systems that cut across the brain-body-world divisions" (Thompson and Varela 2001: 418). According to this view, embodiment means that cognition is not only determined by our bodily perceptual and motor capabilities (Johnson 1987; Lakoff 1987), but also irreducibly constituted by our bodily, multisensory, affective interaction with the environment, in particular by our social interaction (Gallagher 2017: 40–42).

Enactivism explains human social interaction as based on different forms of dynamic sensory-motor couplings that include posture, movement, facial expression, gesture, and linguistic communication (Thompson and Varela 2001: 424; Gallagher 2017: 42). The embodied nature of linguistic communication becomes manifest in co-speech gesture, a human universal feature that develops in early childhood (Iverson and Thelen 1999). Since even speakers with congenital blindness gesture when talking to blind listeners (Iverson and Goldin-Meadow 1998), the association between language and gesture may play some intra-cognitive function. As argued by Kendon (1980, 1995) and McNeill (1992), speech and gesture form a single communicative system based on the same underlying mental processes. These processes are partly externalized in gestures, which therefore offer a window into cognition. However, it is important to note that gesture does not just reflect cognitive processes, but it is itself part of them (McNeill 1992: 245; Goldin-Meadow and Alibali 2013).

Simultaneous interpreters usually gesture in the booth, even when they cannot be seen by the audience (Galhano-Rodrigues 2007; Galhano-Rodrigues and Zagar Galvão 2010; Zagar Galvão 2009, 2013; Adam 2013; Adam and Castro 2013), which suggests that gesture may play some intra-cognitive function in this activity too. Research on co-speech gestures in interpreting can therefore contribute to understanding the embodied cognitive processes of interpreters: the movements of their head, their arms, and their hands may provide valuable information about their strategies of coping with the task *at hand*. This paper presents findings of a descriptive study which aims at exploring the cognitive functions of gesture in simultaneous interpreting. To this end, we have made a detailed, in-depth analysis of a whole process of simultaneous interpreting, focusing on the interpreter's co-speech gestures, mental images, and embodied interaction with her environment. In the following sections we first present the enactive perspective on cognition, focusing on language and gesture (2); we then briefly comment on the literature about gesture in interpreting (3), describe the setting and methodology of our research (4), discuss the results (5), and present some conclusions (6).

2. Enactive cognition

The view of cognition as enacted, embodied, embedded, affective and potentially extended has been gaining currency over the last three decades. Ward and Stapleton (2012) convincingly argued that these five claims about cognition are closely interrelated and interdependent. In particular, the claim that cognition is enactive – i.e., that it depends in an essential way on the activity of the cognizer – entails that it also crucially depends on embodiment and on our relationship to the environment, that it is bound up with affect, and that it can extend beyond the cognizing organism.

2.1 Embodied language

In cognitive linguistics, the origins of the embodiment hypothesis can be traced to cognitive semantics, in particular, to conceptual metaphor theory and the idea that most conceptual metaphors draw on source domains structured by bodily experience (Lakoff and Johnson 1980). This hypothesis was confirmed and further developed by cross-cultural research on metaphor, studies about historical semantic change, and the theory of image schemas (see Rohrer 2006: 124–132).

Another thread of research related to the embodiment of language is the hypothesis of mental simulation, according to which we produce meanings (re)creating embodied mental experiences (Barsalou 1999). This hypothesis means that, in order to understand language, we perform perceptual, motor and affective mental simulations related to the content of utterances (Bergen 2015). Empirical evidence supporting the mental simulation hypothesis comes mainly from behavioral experimentation (see Bergen 2012 for a review), as well as from brain imaging research, studies with brain-damaged patients, and transcranial magnetic stimulation, all of which have demonstrated that, when we perform perceptual or motor mental simulations, we partly use neural circuitry dedicated to perception or action (Kosslyn et al. 2001; Bergen 2015: 143).[1]

A case in point is linguistic memory. Imageability – a word's capability of evoking mental images – has been shown to play a critical role in lexical retrieval (Paivio 1991). The more imageable a word is, the easier it is to remember, since mental images provide an additional path to linguistic memory. This finding is coherent with the hypothesis of mental simulation, and so is the fact that numbers may pose particular difficulty for simultaneous interpreters, which may be related to their lack of imageability (Seeber 2015: 85–86).

1. There is also some brain imaging evidence that indicates independent activation (for a review, see Olivetti Berlardinelli, Palmiero and Di Matteo 2011).

From an enactivist perspective, however, mental simulations should not be understood as inner representations of an outer, pre-given world, but as part of our interaction with the material and cultural environment: "imagination is not something that happens first in the head; it's rather something that involves embodied action, using toys, props, artifacts, instruments, and so on" (Gallagher 2017: 193). In this context, the notion of affordance, taken from ecological psychology (Gibson 1977), plays an important role, since affordances guide our meaningful interaction with the environment. Affordances are environmental combinations of properties perceived as meaningful – as opportunities for action – by an animal. In human face-to-face communication, understanding language involves making use of affordances that include, e. g., gesture, prosody, facial expression, situation and objects present in the environment. From this perspective, imageability could be seen as a particular affordance of words, an opportunity for performing mental simulations that facilitate comprehension.

In relation to linguistic memory, it is interesting to note that, when trying to remember words that do not easily evoke a mental image, people seem to rely more on external support. Even blank locations in space can help us to reduce memory load. In an experiment, Kumcu and Thompson (2020) found that less imageable words triggered more eye movements to the blank locations where verbal information was previously presented. In the retrieval process, part of the memory work was offloaded onto the environment through eye movements to these blank locations. Here, eye movements can be seen as a form of embodied interaction with the spatial locations that afford support to memory.

2.2 Co-speech gesture as embodied cognition

Embodied thinking is visible in gestures, which are an observable aspect of cognitive processes (Goldin-Meadow and Alibali 2013: 269). This is particularly true in the case of co-speech gestures – the idiosyncratic, spontaneous movements of hands and arms that accompany speech –, which provide a window into speakers' perspectives, interests, and mental images. "By looking at the gestures, we can discover, for each person, what was highlighted, what was relevant and what not, and from this infer the imagistic side of their utterances" (McNeill 1992: 110). Let us focus now on some aspects of co-speech gestures that are relevant for our investigation.

2.2.1 Types and functions of gesture

Co-speech gestures have been classified following functional criteria, for example, by McNeill (1992), who distinguished four types: (1) iconic gestures, which depict persons, objects, actions or events; (2) metaphoric gestures, which are also pictorial, but refer to abstract concepts; (3) beats, which are rhythmical hand movements that stress the words they accompany, and (4) deictic gestures, which point to concrete objects and events or to physical loci that metaphorically represent abstract ideas.

Since gestures tend to be multifunctional, these functional categories tend to overlap: for instance, iconic and deictic gestures can be metaphoric, and they can also function as beats. More recent approaches prefer to speak in terms of functional dimensions such as iconicity, indexicality, and metaphoricity (McNeill 2005; Mittelberg and Evola 2014). All these dimensions interact in a given gesture, so that there are no metaphoric or deictic gestures as such, but degrees of metaphoricity or indexicality. In order to establish the local function of gestures, the different dimensions have to be taken into account "in conjunction with the concurrent speech and other contextual factors" (Mittelberg and Evola 2014: 1740).

In this paper we have followed Kendon's (2017) suggestions about the different ways in which gestures "contribute to the meanings of the utterances of which they are a part" (Kendon 2017: 167). From a pragmatic perspective, Kendon (2017) described three main functions of gestures: (1) referential, through which gestures contribute to the referential meaning of the utterance, (2) pragmatic, through which gestures help to organize the structure of the utterance, for example, by providing information about its illocutionary force or by underlining certain parts of the speech, and (3) interactional, through which interaction is regulated. Although there may be overlaps between the referential and pragmatic functions of gestures – for example, a referential gesture may be repeated rhythmically to emphasize a part of speech –, we believe that this classification can provide valuable support in identifying and defining the main local function of a gesture in its context.

2.2.2 Gesture and mental simulation

McNeill (1992) proposed that speech and gesture arise from the same process of utterance formation, which begins with the creation of a mental image and leads to a complex structure integrated by both language and gesture. According to this hypothesis, the imagistic, analogic properties of co-speech gestures emerge from the mental images produced in the initial stages of utterance formation. In a similar vein, the gesture-as-simulated-action framework, developed by Hostetter and Alibali (2008) from an embodied cognition perspective, proposed that gestures emerge from the perceptual and motor mental simulations that underlie language production. They centered their discussion on representational gestures – those that represent the content of speech –, which include iconic, deictic, and metaphoric gestures. These representational gestures were hypothesized to emerge from the activation of motor and perceptual brain areas when mentally simulating actions and perceptions. As McNeill (1992, 2005) suggests, the gesture-as-simulated-action framework "also considers gesture and speech to be two parts of the same cognitive and communicative system" (Hostetter and Alibali 2008: 509), as well as expression of the same mental simulation.

More recent works have provided empirical evidence supporting the hypothesis of gesture as simulated action. For example, Hostetter (2014) showed that speakers make more representational gestures when describing more manipulable objects, and Masson-Carro et al. (2016) found that objects that offer more affordances for action, such as tools, evoke higher representational gesture rates and elicit more gestures imitating actions. It is interesting to note that this effect seems to be restricted to representational gesture, which suggests that this type of gesture may emerge from cognitive processes different from those underlying non-representational gestures, and that each of them may "relate to imagistic and linguistic content in different ways" (Masson-Carro et al. 2016: 437).

2.2.3 Gesture as external support to cognition

The gesture-as-simulated-action framework provides an explanation for how representational gesture emerges from perceptual and motor simulations, but it does not explain the role played by this type of gesture in cognitive processes (Pow et al. 2014). From an enactivist perspective on cognition, co-speech gestures should not be viewed as by-products, but as integral components of the cognitive processes underlying speech production. In fact, there is growing evidence that co-speech gestures fulfill intra-cognitive functions (see Wassenburg et al. 2018 for a review). Embedded/extended approaches to gesture (e.g., Clark 2013; Pow et al. 2014) adopt an enactivist view of the cognitive system as a coupled brain-body-world system that makes use of environmental affordances to distribute cognitive load. According to this view, gestures embody some aspects of a cognitive task and, in doing so, they provide a stable physical and visual presence that externally supports ongoing cognitive processes (Clark 2013). For example, gestures might "allow the gesturer to become aware of structural correlations" (Pow et al. 2014: 10), or they may support the maintenance of a spatial image in memory (Wesp et al. 2001). From an enactivist perspective, gestures do not just emerge from mental simulations; they are also part of these simulations, which cut across the brain-body-world division. In this study, we consider mental images, gestures, and interaction with the environment as different kinds of support for cognition that reside at a certain point of the internal/external continuum.

3. Gestures in simultaneous interpreting

In recent years, interpreting has been described as a multimodal, embodied cognitive activity of inter-lingual mediation that relies on verbal, auditory, visual and motor modalities (e.g., Seeber 2017; Stachowiak-Szymczak 2019). Conference interpreters receive auditory and visual information – that includes speakers' facial expressions and gestures, as well as visual aids like slides and notes –, and combine both types of *input* in a process of audiovisual integration (Seeber 2017).

Moreover, conference interpreters also produce multimodal information. Research on gestures in conference interpreting has revealed that simultaneous interpreters, although not visible to their audience, do gesture in the booth (e. g., Galhano-Rodrigues 2007; Galhano-Rodrigues and Zagar Galvão 2010; Zagar Galvão 2009, 2013; Adam 2013; Adam and Castro 2013). Adam (2013) carried out a detailed study on the role of gestures in simultaneous interpreting, and found that the interpreters gestured more in speech fragments with a higher propositional density, and Chaparro Inzunza (2017) showed that gesturing may have a positive impact on the quality of the interpretation. These results suggest that gestures could provide some kind of cognitive support for interpreters.

Most studies about gesture in simultaneous interpreting reveal that the different types of gestures tend to overlap, which means that gestures tend to be multifunctional. For example, Galhano-Rodrigues (2007) found gestures simultaneously playing structuring, iconic, and deictic functions. She also observed that recurrent gestures play cohesive functions, and that some iconic gestures provide a support to memory, helping the interpreter "to retain meanings while she has not yet verbalized them and is listening to the speaker's next utterances" (Galhano-Rodrigues 2007: 750). Zagar Galvão (2009) found beat gestures with iconic properties, and described an interesting example of close coordination between iconicity and deixis, in which the speaker drew a virtual map of the world in the space in front of him and then pointed out the different areas he was talking about.

Adam and Castro (2013) focused their study on beat gestures and found that they tend to overlap with representative (iconic and metaphoric) gestures. In relation to their specific functions, Adam and Castro's (2013) results suggest that most beat gestures are used for emphasis, aided by prosodic elements; that many of them appear in moments of hesitation, e. g., in connection with self-corrections of the interpreter; and that some of them are cohesive, that is, they mark elements that belong together.

Stachowiak-Szymczak (2019) investigated beat gestures in simultaneous and consecutive interpreting. Her research is based on the idea that interpreting is an embodied activity with a high cognitive load, and that, in order to respond to this load, interpreters resort to bodily activities such as gesture and eye movements, which facilitate cognitive processing. In her experiments, she used numbers and lists as 'problem triggers' (Gile 2009: 192) in order to increase the 'local cognitive load', which she defined as "a temporal demand of dedicated cognitive effort" that corresponds to specific parts of speech (Stachowiak-Szymczak 2019: 46). She found that the mean gaze fixation duration and the number of beat gestures per minute increased when the interpreters had to render numbers and lists, and that the degree of congruence between visual and auditory input influenced how frequently interpreters looked at the screen. On the basis of these findings, Stachowiak-Szymczak (2019: 119) concluded that "beat gestures could be produced in order to better deal with local cognitive effort," and that visual and auditory information is integrated in the interpreting process.

The results obtained so far suggest that interpreters' gestures provide support for their cognitive processes in a number of ways, although it is not yet clear whether different types (or dimensions) of gestures play different roles in the interpreting process. Most of the reported studies focus on just one type of gesture, and most of them analyze just some specific fragments of the interpreting process or use short speeches as input, which makes it difficult to compare the functions of different types of gesture in one interpreting process. In order to provide a broader perspective on the roles played by gestures in simultaneous interpreting, we have investigated a whole interpreting process focusing on the support provided by each type or dimension of gesture.

4. The study

4.1 Objective and research question

The main objective of this exploratory study is to generate hypotheses about the functions of different types of co-speech gestures in simultaneous interpreting. To this end, we have studied the co-speech gestures that emerged throughout a simultaneous interpreting process, along with the concurrent speech and the interpreter's interaction with her environment in the booth, as well as her description of the mental images she produced during the process. This study is based on the idea that language processing is a multimodal activity supported by mental simulation, gesture, and interaction with the environment, and is guided by the following research question: What functions do the different types/dimensions of gestures play in the interpreting process?

4.2 Data collection

The data were collected in May 2018 in an experimental setting at the Universidad de Las Palmas de Gran Canaria, where a remote simultaneous interpreting assignment was simulated and the interpreting processes (English-Spanish) of four participants were separately filmed (Fernández Santana and Martín de León 2021). Participants signed a written consent to anonymously use their data and reproduce their images. For this study we have analyzed the interpreting process of one participant, an English native speaker with 20 years of experience in professional conference interpreting. This participant holds a BA in Modern Languages from the University of Oxford and a PhD in Translation and Interpreting from the Universidad de Las Palmas de Gran Canaria.

The source speech is a TED talk about the transmission of information between bacteria, given by Bonni Bassler in 2009 and entitled *How bacteria "talk*."[2] The initial duration of the video was 17'55''. As the speaker spoke too fast and this could add an undesired difficulty to the interpreting process, we reduced the speed of the video by 10 %, thus increasing its length to 20′ 10″ for the participants to interpret in this experiment. The initial speed reached 191.84 words per minute, and in the edited version for this experiment the final speed was 170.43 words per minute.

Just before entering the booth, the general content of the talk was explained to the participants and they were provided with a short list of specialized terms with their respective translations. It was not until the end of the experiment that the participants were informed of their actual purpose, in order not to interfere with the spontaneity of their gestures. During the interpretation, participants had access to the video through television screens located outside the booths and facing them. The interpreting process was recorded with a Sony HDR-PJ220E camera placed on a tripod, one meter away from the booth, so that it captured a full image of the interpreter from the top of the table. The video resolution is HD 1080.

Immediately after each process, each of the participants was interviewed and the interviews were video recorded. These interviews were based on uncued recall; we formulated open questions about the images, sounds, and movements they remembered having imagined during the interpreting session; about the gestures they remembered having made; about the possible relationships between these gestures and their mental images, and about the possible functions of these gestures in their interpreting process.

2. https://www.ted.com/talks/bonnie_bassler_how_bacteria_talk/transcript

4.3 Analysis of the data

For the identification, analysis and categorization of the gestures, we used the ELAN (Version 6.0) [Computer software]. (2020) (Sloetjes and Wittenburg 2008). For the purpose of identification, we defined gestures as hand movements co-occurring with speech, and discarded self-soothing gestures – such as scratching the head or touching the nose. To quantify gestures, we focused on gesture phrases, which comprise a preparation and a stroke phase and can combine with other gesture phrases into a gesture unit (Kendon 2004; Zagar Galvao 2019).

Following Kendon (2017), we classified the participants' gestures into two main categories: referential and pragmatic, since no interactional gestures were found. Referential gestures contribute to the meaning of the speech, including iconic gestures, which resemble the discourse contents, and deictic or pointing gestures made with a finger, other body part or an artifact. Pragmatic gestures do not allude to the content of the discourse, but rather to its structure and its illocutive force. Pragmatic gestures, according to Kendon (2017: 167–172), include four types: (1) operational, which confirm, negate or deny the content of the speech; (2) modal, which provide a frame for interpreting the speech (e.g., as a quotation, a joke, a hypothesis, etc.); (3) performative, which express the illocutionary force of the accompanying speech (e.g., as a question, an offer, a request, etc.); and (4) parsing or punctuational, which provide emphasis or mark the structure of the speech. In this study, we would identify the interpreter's gestures in terms of the subtypes of the two main categories.

The identification and classification of the gestures was made by the two authors of this paper in different phases. First, each of us worked independently to identify gestures and assign them to one or more categories; then, we discussed doubtful cases and clarified the categories. In a third phase, each author independently assigned each gesture to one main category. To calculate intercoder reliability in this last phase, we used Krippendorff's (1970) alpha coefficient, because, unlike other reliability measures, it is not influenced by factors such as sample size, number of coders, or possible coder bias (Mellinger and Hanson 2017: 211). We obtained an α of 0.982, well over the threshold of 0.80 mentioned by Krippendorf (2004).[3]

We also noted where the interpreter's gaze was directed at each moment, and described the interpreter's interactions with the objects in the booth. These obser-

3. Krippendorff's alpha was computed using R (version 3.6.1) (R Core Team 2017), R Studio (version 1.2.1335, Build 1379 [f1ac3452]) (RStudio Team 2020), 'irr' package (version 0.84.1) (Gamer *et al.* 2019)., and 'lpSolve' package (version 5.6.15) (Berkelaar 2020).

vations were made from the recording, which provided enough detail to determine with certainty whether the participant was looking at the video – which was located at the top left –, or whether she closed her eyes, or looked forward, toward the table, or to the sides.

Finally, we transcribed the interview with the interpreter and categorized the information provided into three main categories: iconic gestures, deictic gestures, and pragmatic gestures. It is worth mentioning that, when talking about her gestures, the participant tended to gesticulate accordingly, which provided additional information about some particular gestures.

The main results of the study are summarized in Section 5.

5. Results and discussion

5.1 Overall picture

In the whole process we identified a total of 358 gestures (see Table 1): 102 referential gestures (52 iconic and 50 deictic), and 256 pragmatic gestures. Among the pragmatic gestures we found a single operational gesture (a negating gesture), 168 performative gestures (including palm up open hands, rotating hands, vertical palm open hands, finger bunches, and praying hands), and 87 parsing gestures (beats and precision grips).

Table 1. Number of gestures found in the whole interpreting process

Function	Types	Subtypes	Number
referential			**102**
	iconic		52
	deictic		50
pragmatic			**256**
	operational	negation	1
	performative		168
		palm up open hands	89
		rotating hands	51
		vertical palm open hands	20
		finger bunches	7
		praying hands	1
	parsing		87
		beats	79
		precision grips	8

As we have seen before, Adam (2013) showed that interpreters gestured more when the propositional density of speech fragments was higher, and Stachowiak-Szymczak (2019) found that the number of beats per minute increased with the cognitive load. A number of studies have shown that, when confronted with difficult cognitive tasks, people switch off from environmental stimuli – including interlocutors and visual displays – closing their eyes or looking at the sky or the floor, so as to concentrate on the cognitive activity at hand (e. g., Glenberg et al. 1998; Kumcu and Thompson 2020). In order to obtain a general picture of the distribution of gestures throughout the interpreting process, we calculated the average number of each type of gesture per minute separately for the fragments in which the interpreter looked at the video and those in which she looked away. Our assumption was that, among the different factors that might lead the interpreter to take her eyes off the screen, one might be the need to concentrate on the interpreting process due to an increased cognitive load, so that the fragments in which she did not look at the video could also show some differences in the density of gestures made per minute.

Table 2. Average number of gestures per minute in the global process and in the fragments where the interpreter looked at the video or looked away

Types	Global	Video	No video
iconic	2.65	2.5	3.9
deictic	2.55	1.97	3.9
referential	5.21	4.02	7.6
negation	0.05	0	0.15
palm up	4.55	5.16	3.28
rotating hands	2.6	3.41	0.93
vertical palm	1.02	1.13	0.78
finger bunch	0.35	0.45	0.15
praying hands	0.05	0	0.15
beat	4.03	4.55	2.96
precision grip	0.4	0.53	0.15
pragmatic	13.08	15.27	8.59
ALL	**18.25**	**19.3**	**16.25**

As can be seen in Table 2 and Figure 2, the average number of gestures per minute was higher in the fragments where the interpreter looked at the video. In terms of density by type of gesture (Figures 1 and 2), the number of pragmatic

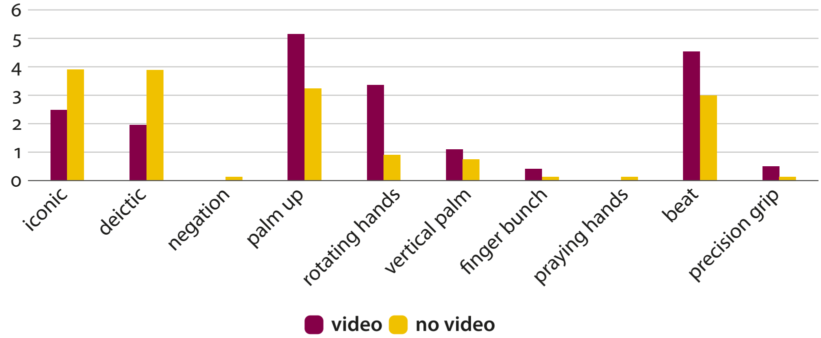

Figure 1. Average number of types of gestures per minute in the fragments where the interpreter looked at the video or looked away

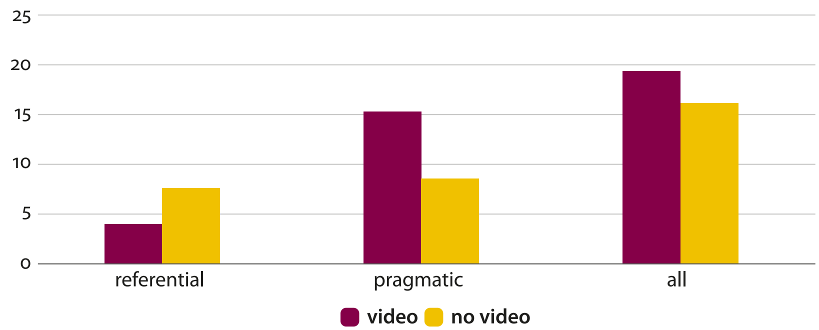

Figure 2. Average number of referential and pragmatic gestures per minute in the fragments where the interpreter looked at the video or looked away

gestures per minute was higher when the interpreter watched the video than when she did not; in contrast, the frequency of referential gestures was higher when she did not watch the screen. Our hypothesis is that, since iconic and deictic gestures tend to co-occur with the production of visual and spatial mental simulations, the images on the screen might interfere with these activities, which would therefore tend to be performed looking away from the video.

In order to obtain more information about the differences between the fragments in which the interpreter looked at the video and those in which she did not, we calculated the ear-voice span at the beginning and end of each fragment. The ear-voice span in simultaneous interpreting is the time lag between the reception of the source speech and the production of the target speech, which reflects cognitive processing, since it can vary from an average of 2 seconds to 10 seconds (Timarová et al. 2011: 121–122). In order to calculate the ear-voice span, we aligned source and interpreted speeches in an Excel file. The alignment was offset based and phrase-level.

As can be observed in Table 3, in the fragments where the interpreter did not look at the screen, the ear-voice span tended to increase with respect to the previous fragment (in 10 out of 14), while in the fragments where she looked at the video it tended to decrease (in 9 out of 13). This pattern was only broken at the beginning and end of the process, and in two other, in-between fragments. The fragments where the interpreter looked away from the video were shorter: in total, they make up just 6 of the approximately 20 minutes of the entire process. In this context, it should be noted that there is a notable difference between an increase or decrease of the ear-voice span in a 3″ fragment and in a 166″ fragment, in which there will be much more oscillations of the time span. The purpose of this quantitative analysis is only to provide an overview of the process in order to develop hypotheses that will need to be tested in future research.

Table 3. Ear-voice span evolution in 27 fragments in which the interpreter alternately looked at the video and looked elsewhere, and length of the fragments

Fragment	No video	Video	Length
1	2″		44″
2		2″ – 5″	255″
3	5″ – 2″		25″
4		2″ – 6″	64″
5	6″ – 13″		18″
6		13″ – 5″	166″
7	5″ – 10″		13″
8		10″ – 7″	102″
9	7″ – 8″		41″
10		8″ – 4″	3″
11	4″ – 8″		16″
12		8″ – 4″	19″
13	4″ – 5″		11″
14		5″ – 4″	23″
15	4″ – 7″		22″
16		7″ – 4″	11″
17	4″ – 2″		29″
18		2″ – 5″	10″

Table 3. *(continued)*

Fragment	No video	Video	Length
19	5″ – 7″		10″
20		7″ – 3″	1″
21	3″ – 8″		12″
22		8″ – 1″	140″
23	1″ – 9″		9″
24		9″ – 3″	4″
25	3″ – 2″		3″
26		2″ – 12″	46″
27	12″ – 6″		17″

The data obtained on the density of referential and pragmatic gestures, as well as on the oscillations of the ear-voice span in the fragments in which the interpreter watched or did not watch the video, provide an approximate overall picture of the interpreting process. This general picture suggests that referential and pragmatic gestures may tend to play different roles in the interpreting process. As Kendon (2017) argued, referential gestures contribute to the meaning of the discourse, whereas pragmatic gestures "indicate type of speech act or aspects of discourse structure" (Kendon 1995: 247), that is, they rather play metadiscursive roles. In our data, referential gestures are more frequent when the interpreter does not watch the video and the ear-voice span tends to increase; these gestures could be helpful when difficulties arise in understanding the source speech content. By contrast, pragmatic gestures, more frequent when the interpreter watches the video and the ear-voice span tends to decrease, could support the progress and structuring of speech. In the following sections we provide a more detailed analysis of some examples of the different types of gestures. These examples have been selected on the basis of their possibility to illustrate the different trends just outlined in the quantitative study. They are also the gestures to which the interpreter mainly referred in the interview conducted after the interpreting process, which suggests that they were particularly salient for her.

5.2 Referential gestures

5.2.1 Iconic gestures: Drawing contents in the air

Some iconic gestures came from imitating the speaker (19 out of 52), although they were not exact replicas. According to McNeill (1992: 133), "iconic gestures, together with the accompanying speech, offer a privileged view of thought. They are the closest look at the ideas of another person that we, the observers, can get." In fact, some of the iconic gestures performed by the interpreter reflect the mental images that she described in the interview, where she explained that she imagined the referents of the speech when she could not find their images on the slides. These mental images were externalized in the form of iconic gestures, most of which were organized in clusters and seemed to support the comprehension of complex concepts such as the symbiosis between squid and bacteria (Example (1)). In the fragments corresponding to these clusters, the ear-voice span tended to increase.

Example 1. Squid and bacteria (6:44 to 7:02)

1. *y cuando tiene* [*exactamente*]$_{\text{beat}}$ [*la misma cantidad de luz*]$_{\text{palm up open hands}}$
   "and when it has [exactly]$_{\text{beat}}$ [the same amount of light] $_{\text{palm up open hands}}$"
2. [*que le llega de la luna o del cielo*]$_{\text{iconic1}}$
   "[coming from the moon or the sky]$_{\text{iconic1}}$

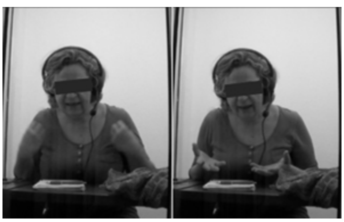

Figure 3. Iconic1

3. [*junto con… eeee…*]$_{\text{iconic2}}$
   "[along with eeee…]$_{\text{iconic2}}$"

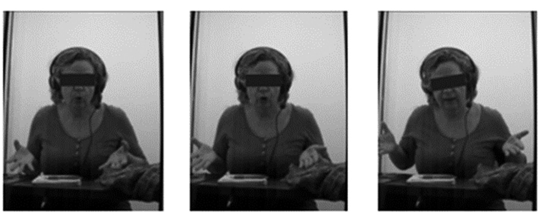

Figure 4. Iconic2

4. [*tiene la misma cantidad de luz*]$_{\text{iconic3}}$
   "[it has the same amount of light]$_{\text{iconic3}}$"

Figure 5. Iconic3

5. [*llegando del cielo*]$_{\text{iconic4}}$ [*que producido por las bacterias*]$_{\text{iconic5}}$
“[coming from the sky]$_{\text{iconic4}}$ [than produced by the bacteria]$_{\text{iconic5}}$”

Figure 6. Iconic4

Figure 7. Iconic5

6. *entonces* [*no hace sombra*]$_{\text{palm up open hands}}$
“then [it doesn’t cast a shadow] $_{\text{palm up open hands}}$”

Here we follow McNeill’s (1992) notation system, and put in brackets the segments of speech that coincide with each gesture. In this fragment about the symbiosis of squid and bacteria, which corresponds to fragment 5 in Table 3, the interpreter looked down, concentrated, and the time lag increased from 6 to 13 seconds. The sequence of gestures contains 1 beat, 2 palm up open hands and 5 iconic gestures that seem to support the interpreter’s effort to understand the idea that the amount of light coming from the sky is the same as that produced by the bacteria inside the squid, so that it does not throw a shadow. At the beginning of the fragment she made a beat and a palm up open hands gesture, then raised both hands towards her shoulders and moved them downwards with fingers extended and palms upwards. This first iconic gesture accompanied *que le llega de la luna o del cielo* and represented the light coming from the sky (Figure 3). Then she attempted to make a gesture in the opposite direction three times, raising her hands palms up, but she lost track of the idea in her speech (Figure 4). Subsequently she repeated the first part (*tiene la misma cantidad de luz*) while she placed her hands up again, with the palms towards her body (Figure 5). Then, while saying *llegando del cielo*, she repeated the top-down gesture (Figure 6), and finally she made the opposite bottom-up gesture accompanying *producido por los bacterias* (Figure 7).

In the interview, the interpreter remembered that she had perfectly imagined the moment when the balance between the two light sources was reached and that this mental image allowed her to understand the symbiotic process. She explained that, since there was no image of this process in the video, she had to imagine it in order to understand it. As stated by Pow et al. (2014: 10), "gesturing might sometimes allow the gesturer to become aware of structural correlations that would be difficult to generate through internal computation." It seems likely that *drawing* this scheme in the air allowed the interpreter to understand the concept, even at the cost of losing a little time and increasing the ear-voice span.

5.2.2 Deictic gestures: Organizing ideas metaphorically in the booth

The most direct use of deictic gestures was to point out the participants in the communication, in this case the audience – imaginary for the interpreter –, and the interpreter herself, as well as the speaker's team when they appeared on the video screen and were the referent of the speech. The interpreter pointed to herself when she referred to the first person in the speech and also, in a metaphorical way, to different referents of the speech (human body, cell, squid, individual). This kind of gestures comprise 12 out of the 50 deictic gestures performed by the interpreter. The rest of them seem to play a more complex role related to the metaphorical use of space.

Pow et al. (2014: 3) observed that "pointing gestures sometimes regulate visuo-spatial attentional processes, being especially helpful under high cognitive task demands," and McNeill (1992: 171) showed that space can be metaphorically dichotomized to represent opposed ideas, which motivates "both abstract pointing – the pointing finger aimed at a concept – and an axial division of space into different roles or meanings." In Example (2), two opposing ideas were placed in opposite places according to the metaphorical structure DIFFERENCE IS DISTANCE.

Example 2. Sorting out ideas (12:36 to 12:39)

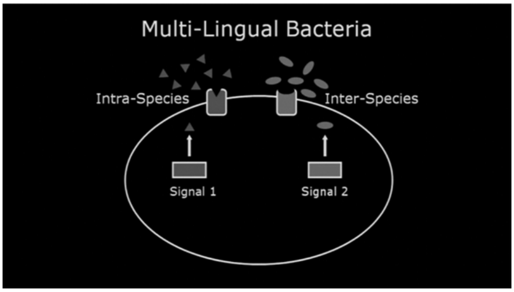

Figure 8. Slide in the source video

Speaker: that bacteria are able [to count]$_{\text{deictic1}}$ [how many of me]$_{\text{deictic2}}$ [and how many of you]$_{\text{deictic3}}$

Figure 9. Deictic1

Figure 10. Deictic2

Figure 11. Deictic3

Interpreter: 7. [*cuántas bacterias*]$_{\text{deictic1}}$ [*de mi especie hay*]$_{\text{deictic2}}$
"[how many bacteria]$_{\text{deictic1}}$ [of my species there are]$_{\text{deictic2}}$"

Figure 12. Deictic1

Figure 13. Deictic2

8. [*y cuántas bacterias*]$_{\text{deictic3}}$ [*de otras especies*]$_{\text{deictic4}}$
"[and how many bacteria]$_{\text{deictic3}}$ [of other species]$_{\text{deictic4}}$"

Figure 14. Deictic3

Figure 15. Deictic4

A few seconds before the beginning of this fragment, the interpreter was watching the video, which showed a slide with the intra-species communication molecules on the left side, and the inter-species communication molecules on the right side (Figure 8). Then, it showed the speaker saying "that bacteria are able to count how many of me and how many of you," while she pointed first to her left side with her left hand (Figures 9–10), and then to her right side with her right hand (Figure 11). Here the interpreter stopped looking at the video and made a chain of 4 deictic gestures (this sequence is part of fragment 11 in Table 3). She first raised her slightly curved right hand and pointed with her fingertips forward, at the level of her head, while saying *cuántas bacterias* (Figure 12); then she pointed to her left, at about the same height, while saying *de mi especie hay* (Figure 13); subsequently she repeated the first gesture saying *y cuántas bacterias* (Figure 14), and finally she pointed to her right while saying *de otras especies* (Figure 15).

She used the space of the booth in front of her to organize the speech and distribute its elements, following the order established by the slide and reproduced by the speaker in her gestures. In the interview, the interpreter referred to the slide and the molecules depicted in it; when she alluded to the small triangles, she moved her hand to the left, and when she referred to the ovals, she moved it to the right, which indicates that after the end of the session she still remembered how the information on the slide was organized.

Example 3. Pointing to the notebook (14:48 to 15:09)

9. *e… esto nos permite hacer* [*moléculas*]$_{\text{deictic1}}$
“th… this allows us to make [molecules]$_{\text{deictic1}}$”

Figure 16. Deictic1

10. *de* [*detección*]$_{\text{deictic2}}$ [*de quórum*]$_{\text{deictic3}}$
“of [quorum]$_{\text{deictic3}}$ [detection]$_{\text{deictic2}}$”

Figure 17. Deictic2

Figure 18. Decitic3

11. [*específicas*]$_{\text{deictic4}}$ [*a determinadas enfermedades*]$_{\text{deictic5}}$
“[specific]$_{\text{deictic4}}$ [to certain diseases]$_{\text{deictic5}}$”
12. [*y aquí los hemos convertido en antagonistas*]$_{\text{deictic6}}$
“[and here we have turned them into antagonists]$_{\text{deictic6}}$”

Figure 19. Deictic6

13. [*al sistema*]$_{\text{deictic7}}$ [*interespecies*]$_{\text{deictic8}}$ [*que esperamos que*]$_{\text{deictic9}}$
“[to the inter-species]$_{\text{deictic7}}$ [system]$_{\text{deictic8}}$ [that we hope]$_{\text{deictic9}}$”

As explained in Section 4.2, before starting the task, the participants were provided with a list of terms with their respective translations, one of which was 'quorum sensing'. The interpreter looked at the notebook when she heard "anti-quorum sensing molecules." She held the pen in her right hand and pointed to different places in the notebook (Deictics 1 to 5 and 7 to 9; Figures 16–18). These gestures also seemed to play the role of beats, as they underlined the rhythm of speech. When she said *y aquí los hemos convertido en antagonistas*, she looked again at the video and drew imaginary circles in the notebook with the tip of the pen (Figure 19). This gesture could also be a metaphorical representation of the antagonistic relationship.

In the interview, the interpreter explained that, although she did not actually draw, when she found the speech more complicated she used the notebook and pen to situate the concepts and then relate them to each other – and by saying this she drew imaginary circles with the pen in the notebook. The first part of this sequence – until the interpreter looked again at the video – corresponds to fragment 19 in Table 2, where the time lag increased from 5 to 7 seconds.

In these last two examples, the booth space and the objects in it provided affordances to the interpreter, who used them mainly to organize ideas metaphorically through deictic gestures. In Example (2), the visual information offered by the slide and the speaker's gestures also supported the interpreter in constructing and organizing the meanings of speech.

5.3 Pragmatic gestures: Managing speech production

As sketched in Section 4.3; Kendon (2017) suggested four functions of pragmatic gestures – operational, modal, performative, and parsing –, all of which can be considered metadiscursive, insofar as they provide information about how the speech has to be understood – for example, when hand actions negate the spoken meaning, or when drawing "quotation marks" in the air. In our data, we have found 1 operational gesture, 0 modal gestures, 168 performative gestures, and 87 parsing gestures. All these gestures seem to fulfill metadiscursive and supportive functions related to speech organization. One of the most frequent performative gestures, with 51 instances, was what we called "rotating hands", a rotating movement of the hands around each other. The frequent use of this gesture – with 2.6 instances per minute – may be idiosyncratic, although rotating gestures have been found to metaphorically represent transitions or processes: "the gesture conveys the transition as repetitive and/or cyclic, an image that appears to be based on rotating wheels or gears, although there are no wheels or gears" (McNeill 1992: 159).

Most of these rotating gestures seemed to be used to manage the interpreter's discursive flow, in particular, when local cognitive load was high and ear-mouth span was big. A recurring pattern consisted of a doubt or a mistake, accompanied by a rotating gesture, and followed by a decision or a repair highlighted by a beat or a palm up open hand. McNeill (1992: 169) described a similar pattern, where "[t]he beat marked the word that was the repair itself, not the word deemed in need of repair."

Example 4. Hesitation (1:56 to 2:00)

Speaker: you are ten percent human, but more likely one percent human

Interpreter: 14. *tiene* [*como uuuun…*]$_{\text{rotating hands}}$ [*uno*]$_{\text{beat}}$ *por ciento de ser humano*

"you have [like aaaaaa…] $_{\text{rotating hands}}$ [one]$_{\text{beat}}$ percent of human being"

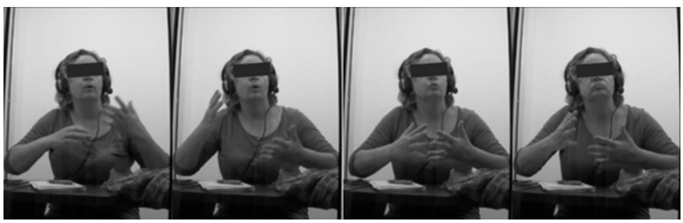

Figure 20. Rotating hands

Figure 21. Beat

In this fragment, the time lag was about 9 seconds, and the interpreter, who was looking at the video, omitted some data. When she made the pragmatic gesture (Figure 20), she hesitated about the percentage and only said the second of those mentioned by the speaker. This decision was underlined by a beat (Figure 21).

Example 5. Repair (5:44 to 5:49)

15. [*y lo que pueden ver aquí son esos dos lóbulas…*]$_{\text{rotating hands}}$
“[and what you can see here are those two *lóbulas…*]$_{\text{rotating hands}}$”

Figure 22. Rotating hands

16. *lo…* [*lóbulos*]$_{\text{beat}}$
“lo… [lobes]$_{\text{beat}}$”

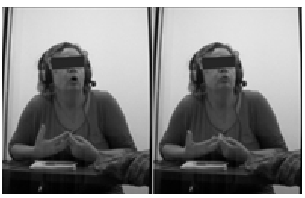

Figure 23. Beat

Here the interpreter was looking at the video too. She made a mistake with the gender of *lóbulos* and then corrected herself. The moment of doubt was accompanied by a rotating gesture (Figure 22); the repair was underlined by a beat (Figure 23).

Example 6. Hedge (2:15 to 2:16)

Speaker: they cover us in an invisible body armor
Interpreter: 17. *nos cubren* [*con una especie de armadura*]$_{\text{rotating hands}}$
“they cover us [with a kind of armour]$_{\text{rotating hands}}$”

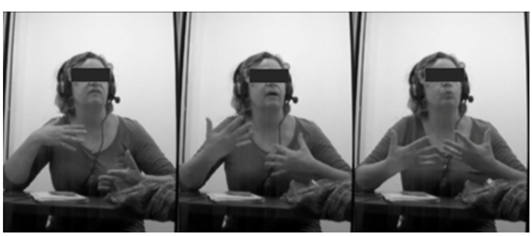

Figure 24. Rotating hands

Rotating hand gestures did not only accompany moments of hesitation and mistakes, they also appeared with metalinguistic comments and hedges – which may also express doubt about a linguistic expression. Here, the rotating hand gesture is in the middle of a chain of iconic gestures that the interpreter made while watching the video (Figure 23). It accompanied the metalinguistic comment on *armadura*, a comment that did not appear in the source speech. In the interview, the interpreter expressed her doubts about this term and, while explaining that she had not been able to find a better term at that moment, she made a rotating gesture with both hands.

Almost half of the rotating hand gestures (22 out of 51) seem to coincide with metalinguistic comments or with moments of hesitation or difficulty in accessing a term, and to help the interpreter to pick up the thread when she has lost it. As they are dynamic gestures, they could support the progression of the interpreting process in moments of doubt or a bottleneck. In fact, 20 of these gestures began with a long ear-voice span and 26 coincided with the omission of information. In general, they are metadiscursive gestures that do not originate from the imitation of the speaker, nor do they refer to the content of the discourse, but to the management of the interpreting process.

During the interview, the interpreter explained that many movements of her hands, more than to create an image, served her to follow the logical thread of what she was saying, while she was listening to a new idea. She referred to this kind of gestures three times along the interview, and all the times she rotated her hands around each other.

6. Conclusions

The results obtained on the frequency of referential and pragmatic gestures in the fragments where the interpreter looked, or did not look, at the video, and on the increasing or decreasing ear-voice span, as well as the information provided by the interpreter in the interview – both orally and gesturally – suggest that referential and pragmatic gestures (Kendon 2017) tend to play different functions in the simultaneous interpreting process:

a. On the one hand, we have found that referential gestures tend to be more frequent in sequences in which the ear-voice span increases while the interpreter seems to be concentrated on understanding the speech content and does not look at the video. In our data, gestures with a high degree of iconicity and highly deictic gestures tend to appear in chains. Some iconic gestures have been found to embody mental images, which is coherent with Hostteter and Alibali's (2008) embodied framework, and most deictic gestures serve to organize information locating ideas metaphorically in the booth space, which is an example of extended cognition. Both activities are related to the content of speech.

b. On the other hand, pragmatic gestures tend to accompany moments of hesitation, mistakes, and metalinguistic comments about the formulation of the discourse (lexicon, morphology), as well as to support the progress of the interpreting process. In these fragments, the interpreter usually watches the video, and the ear-voice span decreases.

These findings allow us to formulate the following hypotheses:

1. The main function of representational gestures in simultaneous interpreting is to support the construction of meanings.
2. Iconic gestures tend to be related to mental images and to support comprehension processes.
3. Deictic gestures help to organize ideas locating them in the booth space.
4. Pragmatic gestures play metadiscursive functions and help to manage the progress of the interpreting process.
5. Simultaneous interpreters make use of the affordances provided by multimodal sources (slides, speaker's gestures, annotations, objects and space in the booth) to understand and organize the contents of the speech.
6. Since iconic and deictic metaphoric gestures tend to co-occur with the production of visual and spatial mental simulations, they tend to be performed looking away from the speaker and from the images that accompany the source speech, which could otherwise interfere with these simulations.

We can conclude that our findings support the idea that simultaneous interpreting is a multimodal, embodied cognitive activity (Seeber 2017; Stachowiak-Szymczak 2019) that can extend beyond the interpreter's organism. One limitation of this work is that it focuses on a single interpreting process, so our findings cannot be generalized. At the same time, however, this is one of its strengths, since we have been able to compare the use of the different types of gestures throughout the whole interpreting process. Future research could focus on testing and refining our hypotheses.

Acknowledgements

We would like to thank two anonymous reviewers for their insightful suggestions and comments, and Matthias Apfelthaler for his valuable help in calculating intercoder reliability.

References

Adam, C. (2013). La relación gesto/complejidad textual en la interpretación simultánea: Un estudio descriptivo. PhD thesis. Universidad de Concepción, Chile.

Adam, C., & Castro, G. (2013). Schlaggesten beim Simultandolmetschen – Auftreten und Funktionen, *Lebende Sprachen 58*(1), 71–82. https://doi.org/10.1515/les-2013-0004

Barsalou, L.W. (1999). Perceptual symbol systems, *Behavioral and Brain Sciences*, *22*, 577–660. https://doi.org/10.1017/S0140525X99002149

Bergen, B. (2012). *Louder than words. The new science of how the mind makes meaning*. New York: Basic Books.

Bergen, B. (2015). Embodiment, simulation and meaning. In N. Riemer (Ed.), *The Routledge handbook of semantics* (pp. 142–157). London: Routledge.

Berkelaar, M. (2020). Interface to 'Lp_solve' v. 5.5 to Solve Linear/Integer Programs. https://github.com/gaborcsardi/lpSolve

Chaparro Inzunza, W.B. (2017). Gesticulación y calidad de la interpretación simultánea – Un estudio experimental. Magister thesis, Universidad de Concepción, Chile.

Clark, A. (2013). Gesture as thought. In Z. Radman (Ed.), *The hand, an organ of the mind: What the manual tells the mental* (pp. 255–268). Cambridge, MA: MIT Press.

ELAN (Version 6.0) [Computer software]. (2020). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Fernández Santana, A., & Martín de León, C. (2021). Con la voz y las manos: Gestos icónicos en interpretación simultánea. With voice and hands: Iconic gestures in simultaneous interpreting. *Hermēneus. Revista de Traducción e Interpretación*, *23*.

Galhano-Rodrigues, I. (2007). Body in interpretation – Nonverbal communication of speaker and interpreter and its relation to words and prosody. In P.A. Schmitt & H. Jüngst (Eds.), *Translationsqualität* (pp. 739–753). Frankfurt: Peter Lang.

Galhano-Rodrigues, I., & Zagar Galvão, E. (2010). The importance of listening with one's eyes: A case study of multimodality in simultaneous interpreting. In J. Díaz Cintas, A. Matamala & J. Neves (Eds.), *New insights into audiovisual translation and media accessibility* (pp. 241–253). Amsterdam: Rodopi.

Gallagher, S. (2017). *Enactivist interventions. rethinking the mind*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198794325.001.0001

Gamer, M., Lemon, J. & Fellows Puspendra Singh, I. (2019). Package 'irr', version 0.84.1. https://www.r-project.org

Gibson, J.J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing* (pp. 67–82). Hilldale, NJ: Erlbaum.

Gile, D. (2009). *Basic concepts and models for interpreter and translator training (revised edition)*. Amsterdam: John Benjamins. https://doi.org/10.1075/btl.8

Glenberg, A.M., Schroeder, J.L., & Robertson, D.A. (1998). Averting the gaze disengages the environment and facilitates remembering. *Memory & Cognition 26*(4), 651–658. https://doi.org/10.3758/BF03211385

Goldin-Meadow, S., & Alibali, M.W. (2013). Gesture's Role in Speaking, Learning, and Creating Language. *Annual Review of Psychology 64*, 257–283. https://doi.org/10.1146/annurev-psych-113011-143802

Hostetter, A.B. (2014). Action attenuates the effect of visibility on gesture rates. *Cognitive Science*, *38*(7): 1468–1481. https://doi.org/10.1111/cogs.12113

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495–514. https://doi.org/10.3758/PBR.15.3.495

Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature 396*, 228. https://doi.org/10.1038/24300

Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, *6*(11–12), 19–40.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226177847.001.0001

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The relation between verbal and nonverbal communication* (pp. 207–227). The Hague: Mouton. https://doi.org/10.1515/9783110813098.207

Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of Pragmatics*, *23*, 247–279. https://doi.org/10.1016/0378-2166(94)00037-F

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: 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

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, *2*, 635–642. https://doi.org/10.1038/35090055

Krippendorff, K. (1970). Estimating the reliability, systematic error, and random error of interval data. *Educational and Psychological Measurement*, *30*(2), 61–70. https://doi.org/10.1177/001316447003000105

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: SAGE Publications.

Kumcu, A., & Thompson, L. R. (2020). Less imageable words lead to more looks to blank locations during memory retrieval. *Psychological Research*, *84*(3), 667–684. https://doi.org/10.1007/s00426-018-1084-6

Lakoff, G. (1987). *Women, fire, and dangerous things: what categories reveal about the mind*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226471013.001.0001

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Masson-Carro, I., Goudbeek, M., & Krahmer, E. (2016). Can you handle this? The impact of object affordances on how co-speech gestures are produced. *Language, Cognition and Neuroscience*, *31*(3), 430–440. https://doi.org/10.1080/23273798.2015.1108448

McNeill, D. (1992). *Hand and mind: what gestures reveal about thought*. Chicago: University of Chicago Press.

McNeill, D. (2005). *Gesture and thought*. Chicago: Chicago University Press. https://doi.org/10.7208/chicago/9780226514642.001.0001

Mellinger, C. D., & Hanson, T. A. (2017). *Quantitative research methods in translation and interpreting studies*. London/New York: Routledge.

Mittelberg, I., & Evola, V. (2014). Iconic and representational gestures. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill & J. Bressem (Eds.), *Body – language – communication: An international handbook on multimodality in human interaction* (pp. 1732–1746). Berlin/Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110302028.1732

Olivetti Belardinelli, M., Palmiero, M., & Di Matteo, R. (2011). How fMRI technology contributes to the advancement of research in mental imagery: A review. In J. F. P. Peres (Ed.), *Neuroimaging for clinicians – Combining research and practise* (pp. 320–346). London: IntechOpen. https://doi.org/10.5772/25051

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *45*(3), 255–287. https://doi.org/10.1037/h0084295

Pouw, W. T. J. L., de Nooijer, J. A., van Gog, T., Zwaan, R. A., & Paas, F. (2014). Toward a more embedded/extended perspective on the cognitive function of gestures. *Frontiers in Psychology*, *5*, Article 359. https://doi.org/10.3389/fpsyg.2014.00359

R Core Team. (2017). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org/

Rohrer, T. (2006). Three dogmas of embodiment: Cognitive linguistics as a cognitive science. In G. Kristiansen, M. Achard, R. Dirven & F. J. Ruiz de Mendoza Ibáñez (Eds.), *Cognitive Linguistics: Current applications and future perspectives* (pp. 119–146). Berlin, New York: Mouton de Gruyter.

RStudio Team. (2020). *RStudio: Integrated Development for R.* RStudio, PBC, Boston, MA. http://www.rstudio.com/

Seeber, K. (2015). Simultaneous interpreting. In H. Mikkelson & R. Jourdenais (Eds.), *The Routledge handbook of interpreting* (pp. 79–95). London: Routledge.

Seeber, K. (2017). Multimodal processing in simultaneous interpreting. In J. W. Schwieter & A. Ferreira (Eds.), *The handbook of translation and cognition* (pp. 461–475). Howoken, NJ: Wiley Blackwell. https://doi.org/10.1002/9781119241485.ch25

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)* (pp. 816–820).

Stachowiak-Szymczak, K. (2019). *Eye movements and gestures in simultaneous and consecutive interpreting*. Cham: Springer. https://doi.org/10.1007/978-3-030-19443-7

Thompson, E. & Varela, F. (2001). Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences*, *5*(10), 418–425. https://doi.org/10.1016/S1364-6613(00)01750-2

Timarová, Š., Dragsted, B., & Hansen, I. G. (2011). Time lag in translation and interpreting. In C. Alvstad, A. Hild, & E. Tiselius (Eds.), *Methods and strategies of process research: Integrative approaches in Translation Studies* (pp. 121–146). Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/btl.94.10tim

Ward, D., & Stapleton, M. (2012). Es are good. Cognition as enacted, embodied, embedded, affective and extended. In F. Paglieri (Ed.), *Consciousness in interaction: The role of the natural and social context in shaping consciousness* (pp. 89–104). Philadelphia: Benjamins. https://doi.org/10.1075/aicr.86.06war

Wassenburg, S. I., de Koning, B. B., & van der Schooot, M. (2018). In which direction to move? Facilitative and interference effects of gestures on problem solvers' thinking. *Journal of Cognitive Psychology*, *30*(3), 307–313. https://doi.org/10.1080/20445911.2018.1432628

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *American Journal of Psychology*, *114*, 591–600. https://doi.org/10.2307/1423612

Zagar Galvão, E. (2009). Speech and Gesture in the Booth – A Descriptive Approach to Multimodality in Simultaneous Interpreting. In D. de Crom (Ed.), *Selected Papers of the CETRA Research Seminar in Translation Studies 2008*. Retrieved from <https://bit.ly/2Ik9t4k>

Zagar Galvão, E. (2013). Hand gestures and speech production in the booth: do simultaneous interpreters imitate the speaker? In C. Carapinha & I.A. Santos (Eds.), *Estudos de linguística*, vol *II* (pp. 115–129). Coimbra: Imprensa da Universidade de Coimbra.
Zagar Galvão, E. (2019). Gesture functions and gestural style in simultaneous interpreting. In H. Salaets & G. Brône (Eds.), *Linking up with video: Perspectives on interpreting practice and research* (pp. 151–179). Amsterdam: John Benjamins.

Address for correspondence

Celia Martín de León
Universidad de Las Palmas de Gran Canaria
c/Pérez del Toro 1
35003 Las Palmas de Gran Canaria
celia.martin@ulpgc.es

Biographical notes

Celia Martín de León is Associate Professor of Translation at the Universidad de Las Palmas de Gran Canaria, Spain. Her main research interests are the analysis of conceptual metaphors and implicit theories about translation, and the study of translation and interpreting processes from the viewpoint of embodied and embedded cognition, with special emphasis on meaning construction and imagination. She has authored several articles and book contributions on these topics.

https://orcid.org/0000-0003-4548-9033

Alba Fernández Santana is a Postgraduate Alumni of Translation and Interpreting with a Master's degree in Professional Translation, Intercultural Mediation and Interpreting at the Universidad de Las Palmas de Gran Canaria, Spain. She has focused her research projects on the study of simultaneous interpreters' gestures in the booth focusing on meaning construction, concept retrieval and other cognitive processes that underlie speech production and simultaneous interpreting.

https://orcid.org/0000-0002-7330-1483

Publication history

Date received: 30 September 2020
Date accepted: 5 April 2021