Journal of Psycholinguistic Research

Observations on the progression of gestures with L2 proficiency: A call for further research.

--Manuscript Draft--

Manuscript Number:	JOPR-D-21-00012R2
Full Title:	Observations on the progression of gestures with L2 proficiency: A call for further research.
Article Type:	Original Research
Keywords:	Gesture; second Language Acquisition; proficiency; Mental models; Chunking; Pragmatic gestures
Corresponding Author:	Renia Lopez-Ozieblo Hong Kong Polytechnic University Kowloon, HK HONG KONG
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Hong Kong Polytechnic University
Corresponding Author's Secondary Institution:	
First Author:	Renia Lopez-Ozieblo
First Author Secondary Information:	
Order of Authors:	Renia Lopez-Ozieblo
Order of Authors Secondary Information:	
Funding Information:	Hong Kong Polytechnic University (P00009723) — Dr. Renia Lopez-Ozieblo
Abstract:	Gestures in learners of second languages (L2) are still an understudied field. Most studies comparing the same individuals gesturing in their L2 and in the mother tongue (L1) report a higher frequency of gestures in the L2, however, the reason for this difference is not yet clear. Aside from idiosyncratic and cultural variations, differences in language type, such as syllabic duration and functional load and syntactical structure, as well as differences in the proficiency of the speaker are likely to affect the production of gestures in terms of their frequency and function. The many variables to be considered obscure the role of proficiency and the integration of gesture-speech-idea in L2 speakers. These variables are only controlled in a handful of longitudinal studies comparing speakers' gestures at two or more points in the proficiency continuum. With this paper, we seek to add our observations to this corpus, using the data collected under the case study of a Cantonese learner of English, to reflect on a number of existing theoretical constructs. We propose a number of ideas that need further research, as well as a number of variables that need to be better understood in order to advance L2 gesture research. As our observations are based on the narrations of one individual, our generalizations are theoretical at this point as our results cannot to be extrapolated to all L2 learners, instead, they are intended to highlight a number of areas related to L2 gesture production and how this might change with proficiency, that require further study.

Title Page with ALL Author Affiliation

Author

Renia Lopez-Ozieblo

Title

Observations on the progression of gestures with L2 proficiency: A call for further research.

Affiliation

Department of English and Communication

The Hong Kong Polytechnic University

Tower A, 4th Floor,

Hung Hom, Hong Kong

Email

Renia.lopez@polyu.edu.hk

ORCID

0000-0003-1752-9392

Abstract

Gestures in learners of second languages (L2) are still an understudied field. Most studies comparing the same individuals gesturing in their L2 and in the mother tongue (L1) report a higher frequency of gestures in the L2, however, the reason for this difference is not yet clear. Aside from idiosyncratic and cultural variations, differences in language type, such as syllabic duration and functional load and syntactical structure, as well as differences in the proficiency of the speaker are likely to affect the production of gestures in terms of their frequency and function. The many variables to be considered obscure the role of proficiency and the integration of gesture-speech-idea in L2 speakers. These variables are only controlled in a handful of longitudinal studies comparing speakers' gestures at two or more points in the proficiency continuum. With this paper, we seek to add our observations to this corpus, using the data collected under the case study of a Cantonese learner of English, to reflect on a number of existing theoretical constructs. We propose a number of ideas that need further research, as well as a number of variables that need to be better understood in order to advance L2 gesture research. As our observations are based on the narrations of one individual, our generalizations are theoretical at this point as our results cannot to be extrapolated to all L2 learners, instead, they are intended to highlight a number of areas related to L2 gesture production and how this might change with proficiency, that require further study.

Keywords

Gesture, Second Language Acquisition, Proficiency, Mental models, Chunking

Acknowledgements

Thank you to P for his participation.

This study was funded by the The Hong Kong Polytechnic University, Central Research Grant for projects considered by the Hong Kong University Grant Committee, project number P00009723. The PI of the project is a member of the Research Centre for Professional Communication in English (RCPCE) of the Department of English and Communication of the Hong Kong Polytechnic University, the mission of which is to pursue applied research and consultancy to deepen the understanding of professional communication and to better serve the communicative needs of professional communities. This project is intended to fulfil in part its mission.

Declarations

Funding: funded by The Hong Kong Polytechnic University, Central Research Grant for projects considered by the Hong Kong University Grant Committee, project number P00009723.

Conflicts of interest/Competing interests (include appropriate disclosures): None

Ethics approval (include appropriate approvals or waivers): Approval obtained from the Human Subjects Ethics Sub-committee of the Hong Kong Polytechnic University, application number HSEARS20170227008.

Consent to participate (include appropriate statements): Consent was obtained from the participant.

Consent for publication (include appropriate statements): Consent was obtained from the participant.

Availability of data and material (data transparency): The transcribed data will be made available in a public repository. The raw data (videos) can only be shared upon request, as it will require further consent from the participant.

Code availability (software application or custom code): N/A

Observations on the progression of gestures with L2 proficiency: A call for further research.

Abstract

Gestures in learners of second languages (L2) are still an understudied field. Most studies comparing the same individuals gesturing in their L2 and in the mother tongue (L1) report a higher frequency of gestures in the L2, however, the reason for this difference is not yet clear. Aside from idiosyncratic and cultural variations, differences in language type, such as syllabic duration and functional load and syntactical structure, as well as differences in the proficiency of the speaker are likely to affect the production of gestures in terms of their frequency and function. The many variables to be considered obscure the role of proficiency and the integration of gesture-speech-idea in L2 speakers. These variables are only controlled in a handful of longitudinal studies comparing speakers' gestures at two or more points in the proficiency continuum. With this paper, we seek to add our observations to this corpus, using the data collected under the case study of a Cantonese learner of English, to reflect on a number of existing theoretical constructs. We propose a number of ideas that need further research, as well as a number of variables that need to be better understood in order to advance L2 gesture research. As our observations are based on the narrations of one individual, our generalizations are theoretical at this point as our results cannot to be extrapolated to all L2 learners, instead, they are intended to highlight a number of areas related to L2 gesture production and how this might change with proficiency, that require further study.

1. Introduction

Gestures, meaning-making hand and arm movements co-occurring with speech in second language (L2) learners, provide valuable information on the "communicative and psychological aspects of [L2] development" (Gullberg & McCafferty, 2008, p. 136). However, it is still not fully understood how gestures develop in language learners (Brown & Gullberg, 2008), what their functions are and how they are processed.

Existing gesture studies often focus on referential gestures and only provide a static snapshot of gestures at a specific proficiency point. With this cross-sectional information, it is not possible to understand how students are

progressing in processing the L2. A number of previous studies have compared referential gestures in bilinguals in their two languages, noting that bilinguals tend to gesture more than monolinguals (Nicoladis el al., 2007; Nicoladis, Pika & Marentette, 2009; Pika, Nicoladis & Marentette, 2006) and highlighting cross-cultural variations (Kita 2009; Pika et al., 2006). However, there are few studies on how gestures in general, not just referential, develop within adult language learners in the L2, although those comparing speakers of different proficiency levels suggest that gesture rates do vary by proficiency and type of gestures (Lin, 2019; Sherman & Nicoladis, 2004; Zhao, 2006). Due to the idiosyncratic nature of gestures, the effect of proficiency can only be answered by longitudinal studies following individuals as they improve their L2, which only a handful of studies have addressed so far (Saddour, 2017; Stam, 2015).

Our study provides additional data on how gestures develop with proficiency in one individual. We observed some unexpected behavior in gestures with pragmatic functions, which increased with proficiency. These include beats but also a number of recurrent gestures which can be observed in both the L1 and the L2 (Author, 2020). Based on these observations we propose a number of (theoretical) ideas that deserve further study. The data is extracted from a quantitative analysis of the gestures of a Cantonese speaker narrating a cartoon in his L1 and in his second language (English) at different times, his proficiency in the L2 having increased in the interval. The study reports on the frequency of different types of gestures, their duration and semantic synchronicity with their affiliated lexical unit. The overall hypothesis is that the production of gestures changes with the development of L2 proficiency in adult language learners. Although the limited corpus does not allow us to prove this hypothesis, our observations have led us to consider a number of existing theoretical constructs and suggest practical implications that may advance research into L2 gestures.

2. Mental models

Grounded cognition theories advocate that speakers' conceptualizations take the form of dynamic mental models which are integrated with the linguistic system (Barsalou et al., 2008; Cutica & Bucciarelli, 2015). The more detailed these models are, the faster and more accurate the externalisation of the thought should be (Cutica & Bucciarelli, 2015). At the same time, the more developed the skills to transfer the model's representations into words (automation linked to proficiency), the more streamlined the message ought to be, following Grice's maxim of quantity (1975). In

effective communicative acts, gesture and speech complement each other to make meaning, externalizing and pushing out speakers' conceptualizations or ideas (McNeill, 2005).

In native speakers, the function of the gesture is considered to be dual: communicative (for the benefit of the interlocutor) as well as cognitive (to aid the speaker) (Hadar, 2014; Kita, et al., 2017; Saddour, 2017). Hypotheses as to how gesture is produced support either a modular approach (Hadar, 2014, based on Levelt's 1993, speech model) or a more integrated one, with the idea, speech and gesture conceptualized as a unit (McNeill, 2005). However, both hypotheses defend an automated system. In L2 speakers, this automation might not be fully developed, as is the case with speech (Kormos, 2014). Observing how gestures change with proficiency in L2 speakers will enhance our understanding of how gestures are processed at different stages of proficiency and how their functions develop and provide some insight into how the dynamic mental model is integrated with the linguistic system.

3. Gesture

Gestures can be categorized by form and their relationship with the semantic or pragmatic content of speech. The former are 'referential' – re-enacting, drawing, moulding, representing or pointing to the speech content (Müller, 1998). Pointing gestures, also known as deictic, help establish connections between units within the utterance functioning as anaphoric references. Gestures can also have pragmatic functions: metadiscursive, interactive or cognitive. Metadiscursive gestures help manage the flow of the discourse, making parts of it salient, like beats, indicating clarifications or disfluencies. Interactive gestures engage interlocutors by pointing at them, managing the turn or ensuring understanding. Cognitive gestures link parts of the utterance by indicating inferences, causality and consequence or they provide additional modal information about the speaker's stance on the utterance (Author, 2020). These gestures help the flow of the discourse and include recurrent gestures of a metaphorical nature which can be observed in different speakers and contexts, such as turning the wrist to open the palm upwards (Müller, 2004).

Existing studies seldom categorize gesture by function, focusing instead on its relation to the speech, following Kendon's Continuum as proposed by McNeill (1992). According to this, gestures can be identified by their nature as iconic, deictic, beats or metaphoric, although in many cases the same gesture might display more than one of these characteristics. Metaphoric gestures in particular can be difficult to categorize, due to their pragmatic functionality,

and are reported inconsistently across studies. For example, Lin (2019) describes metaphorical gestures as representing "a metaphor of the speaker's idea or feeling about a concrete concept" (2019, p. 6) while Andric and Small refer to them as gestures that "provide semantic meaning" (2012, p. 99). Other studies report instead on gestures with pragmatic functions referring to them as affect displays (Zhao, 2006), which are "gestures that express emotion" or regulators, gestures to manage turn-taking (Gregersen et al., 2009, p. 199).

In this study we propose a categorization based on the functionality of the gestures (Kok et al., 2016; Author, 2020) believing that it might give a clearer indication as to what exactly is changing with proficiency. Gestures with semantic and/or pragmatic functions can complement the speech, contradict it (a mismatch) or be redundant to it. The frequency and type of gestures, categorized by their form and relationship to the semantic content of speech have been analysed extensively in L2 speakers of various languages, generally concluding that there are more gestures in the L2 than in the L1 (Stam, 2006; Lin, 2019; Zhao, 2006), but their functions are still not fully understood. Existing studies with high proficiency L2 speakers suggest a similar rate of referent (iconic) gestures in the L2 and the L1 in both interactions and narrations (Lin, 2019; Sherman & Nicoladis, 2004 respectively). However, in lower proficiency speakers, there seem to be more iconic gestures in the L1 than the L2, at least in interactions (Lin, 2019; Zhao, 2006). Negueruela and Lantolf (2009), in a study comparing speakers of both English and Spanish as their L1 or L2, concluded that in the L1 referential gestures were more likely to be communicative, adding complementary meaning, while in the L2 they were more likely to be self-regulatory and duplicated the speech content. This supports McCafferty's (2009) argument that in the L2 gestures might facilitate speech production in lower proficiency speakers and Gullberg's conclusions (2011) that low proficiency speakers use referential gestures to elicit help from their listeners. Deictic gestures have been observed to decrease with proficiency, related to an increase in anaphoric structures (Gullberg, 2006), or as a compensatory strategy to specify a physically present referent (So et al., 2013) or an abstract one (Lin, 2019).

Although gestures with pragmatic functions occur in L2 speech, except for beats (Im & Baumann, 2020; Gluhareva & Prieto, 2017; Igualada et al., 2017; Kushch et al., 2018; Rohrer et al., 2020; Shattuck-Hufnagel & Prieto, 2019), they have not received much attention from researchers. Speakers of lower proficiency have been noted to produce more beats than more proficient speakers (McCafferty, 2004, 2006; Zhao, 2006). Beats are linked to prosody, allowing

speakers to mark salience (Rohrer et al., 2020), stance (Shattuck-Hufnagel & Prieto, 2019) or structure (Im & Baumann, 2020) – a function often observed in lower proficiency speakers (McCafferty, 2006). It is likely that lower proficiency L2 speakers might be using gestures instead of linguistic devices for a range of pragmatic functions such as to keep the floor, stress salient information, keep the structure of the utterance or link sequences, among others (Graziano, 2020). We know even less about these functions in the L2 speaker and how the gesture might aid them, than we do about their use of referential gestures.

3.1 Cantonese gestures

Speakers of different languages conceptualize differently (Gullberg, 2009; Slobin, 2006). This is most obvious in mismatches between referential gestures and their lexical affiliate, e.g. the verb encodes path information but the gesture indicates manner (Gullberg, 2009; Stam, 2006, 2015). The few studies available on the gestures of Cantonese speakers suggest that structural differences between English and Cantonese (Slobin, 2006) might also be reflected in how path and manner are expressed in gestures. In English, a satellite-framed language, verbs usually encode manner while path is encoded in another unit, the satellite. Both Mandarin and Cantonese are considered equipollently-framed languages where manner is encoded in a first-occurring verb and path in a second (Talmy, 2000). Motion information encoded in the gesture of Chinese (Mandarin) speakers is predominantly the path, while in English speakers it is path and manner (Kita & Özyürek, 2003). These potential differences could create mismatches in L2 speakers between the speech and the gesture which are expected to be resolved as proficiency develops.

Cantonese is described as having a topic-comment structure, where "the topic is a phrase the comment says something about" (Matthews & Yip, 2011, p. 77). Word order is relevant to express specific grammatical relations between the various parts of the sentence, usually following a subject + verb + object structure. However, deviations to this structure occur in specific cases (see Matthews & Yip, 2011, for a detailed explanation). Particles, some final, are added to indicate aspect, mood and even meaning.

Synchronicity gesture-speech in Cantonese might be more variable than that observed in other languages, gesture being equally likely to be synchronous, precede or follow speech (Chui, 2005). Fung and Mok (nd) observed a lack of prosody-gesture alignment in Cantonese, unlike that observed in English (McNeill, 2005). It has been reported that

both Cantonese and Mandarin speakers might have gesture rates lower than English speakers (Kwan et al., 2012; So, 2010 respectively) but it is not clear whether Cantonese-Mandarin gesture rates are equivalent. Gestures in Chinese speakers, and in particular in Cantonese speakers, are still understudied (Hou & So, 2014). Thus, this work also provides much needed-data to the field. This study explores some of the gaps in the literature concerning our knowledge of Cantonese (L1) - English (L2) speakers. In particular we theorise as to how gestures might be developing with proficiency, in this one Cantonese speaker learning English, and how they compare with the gestures observed in his L1 speech. We provide a detailed analysis of this speakers' gestures leading to a number of observations regarding their functions and processing.

4. Research Questions

The main question we sought to answer was whether there was a change in gesture frequency or function with proficiency in one individual whose English proficiency had improved during a period of six months. An additional related question sought to identify what other variables, aside from proficiency, might be affecting his gesture production.

5. The Study

This study followed a quantitative analysis of the frequency and duration of gestures in a corpus of narrations from one participant, P. P, together with other participants, had answered a call to volunteer in a gesture-related study that sought to identify differences in gestures in English speakers with different proficiency levels. He agreed to repeat the session, allowing us to track changes in his gestures as his proficiency improved. At the beginning of the study P was a 27 year old student in his third year at the XXX university in XXX where the medium of instruction is English. P is a Hong Kong native whose L1 and home environment is Cantonese. He started learning English as a second language at primary school and continued in secondary school; both institutions used Cantonese as the medium of instruction. After finishing school, he worked in a Cantonese speaking environment, and it was only upon starting university that he began to use English regularly. Both English and Cantonese are official languages in Hong Kong. Taking Grosjean's definition of bilinguals as "people who need and use two or more languages (or dialects) in their everyday lives" (2010, p. 4), P is an English/Cantonese bilingual speaker. However, he believes himself to be a learner of

English and views English as his second language. We understand bilingualism is a continuum, however, for the sake of clarity, we will make a distinction between fully bilingual individuals, those speakers "who can use both languages with equal or near equal fluency" (as per the definition given by the Collins English Dictionary online), and those who are still learning the language (L2 learners).

At the time of the study, P's proficiency was evaluated to be between Upper Intermediate and Advanced (B2 and C1 in The Common European Framework of Reference for Languages, CEFR). The narrations included one in Cantonese, P's L1 and two in English. These were recorded six months apart, a period over which P reported that his English had improved, which was confirmed by two independent evaluators. P volunteered to participate in the study and also helped to confirm some of the transcriptions.

5.1 Session Procedure

Sessions lasted an hour (the first one) and 30 minutes (the second one, although there was also an additional 30 minutes of informal talk, before and after the session). For the first session, after an informal chat with the participant to collect some personal details and further explain the project, P was given a short story to read, as many times as necessary to be able to narrate the story to the researcher without referring to the text. The process was repeated with an audio task. These tasks were carried out to familiarize P with the process and reduce anxiety. Then, P viewed the first half of the 'Tweety and Sylvester' episode, Canary Row, (Freleng, 1950). This replicates well-tested input used in many studies (Brown & Gullberg, 2008; Kita & Özyürek, 2003; McNeill, 2005; Stam, 2006). P was allowed to view the video as many times as he wanted, allowing him to prepare to narrate the story in Cantonese (preparation has been found to decrease the number of gestures produced during a narration, Lin 2020). After the narration in Cantonese, P was asked to watch the second half of the video, again, as many times as he felt necessary, and then narrate the story in English. The researcher, known to P, was the interlocutor and the session was recorded with P's consent, but he was not aware the study focused on gestures. The same task, with a different story was repeated after six months during which P felt his proficiency in English had improved. The videos were watched without sound and the extracts selected were of a similar length and contained a similar number of characters and events.

5.2 Analytical Procedure

P's oral proficiency was evaluated with a series of analytical parameters, and also by two independent proficiency evaluators, blind to the sequence of the narrations, using the recordings and sharing rubrics. Analytical parameters have been found to provide an accurate assessment of the three elements of oral proficiency: complexity, accuracy and fluency. The analytical parameters evaluated in this study were: vocabulary (token and type) as a measure of accuracy and complexity; speech rate (words per minute), clauses per minute and cohesive markers per minute to measure structural fluency; and words per clause to measure complexity (Bernstein et al., 2010; Iwashita et al., 2008; Kormos, 2006). The two evaluators rated P.'s proficiency on a scale of 1 to 100%, according to a rubric adapted from the International English Language Testing System rubric, common in Hong Kong, itself based on the guidelines of the Common European Framework of Reference for Languages. They evaluated coherence, diction, literary devices, grammar, accuracy – based on the input – pronunciation, transition, intonation, parallelism and fluency. The evaluators used the audio of the narrations only, not the full video recording, in order not to be influenced by P.'s gestures, as research indicates that native speakers are influenced by speakers' gestures when it comes to assessing their oral proficiency (Billot-Vasquez et al., 2020; Neu, 1990). All the evaluations indicated there had been an improvement in L2 proficiency (from an average of 65% to 69.5%, according to the evaluators' scores).

Each session was video-recorded and the speech and gestures in the recordings – both the Cantonese (Ct) and English narrations (En1) of the first session and the English narration (En2) of the second session – were transcribed using *PRAAT* (Boersma & Weenink, 2019) (a speech transcription software). Gestures were further coded in *ELAN* (Sloetjes, 2017) (a transcription software for multimodal data) by two independent coders and then checked by a third coder. Both the speech and gesture transcription were adapted from McNeill (2005). Gesture classification was simplified into semantic gestures, either referential or deictic, and gestures with a pragmatic function. These latter ones were subdivided into beats (with metadiscursive functions) and other discursive gestures (with metadiscursive, interactive or cognitive functions). Beats were considered separately as there has been some indication of their use in lower proficiencies to mark the structure of the utterance (McCafferty, 2006). In gesture chains, each gesture was counted separately, unless they were beats (a succession of quick up and down movements with no other changes in form or speed). The gesture was divided into phases: the stroke is considered the meaning-carrying part of the gesture,

often, but not always, preceded by a preparation movement and followed by a return to the rest position with or without holds in between the phases. Gesture duration includes all the phases, from rest to rest position. The part of the speech that was synchronous with any part of the stroke was noted as the related lexical unit (even if not always its lexical affiliate).

WordSmith (Scott, 2004) (a Corpus analysis software) was used to calculate the number of words per narration for the English narrations. These and the number of clauses and events narrated were also manually counted by two English speaking researchers and by two Cantonese speaking researchers for the Cantonese narration, reaching 100% agreement. To assess proficiency changes we also used VocabProfile (Cobb, 2002; Heatley et al., 2002) to identify number and type of tokens, including cohesive linguistic units.

From the *ELAN* files we extracted the lengths of the narrations, lengths of gestures, types of gestures, speech pauses and gesture holds, and the lexical affiliate with which the gestures occurred. Durations are given in ELAN in milliseconds, the inter-rater reliability for the observations on duration was high, a Cronbach alpha of 0.94 was calculated on a sample of 20% of the data. For observations on the frequency of gestures per type, the final inter-rater reliability Cronbach alpha was 0.97 on the full sample. The data was then analysed in *JASP* (JASP Team 2020, a statistical analysis software). A number of paired samples Wilcoxon signed-rank tests were performed to compare the quantitative data on gesture duration from the three narrations. A non-parametric test was chosen as not all of the data pairs complied with the Shapiro-Wilk normality check (in some cases there were not many data points). Repeated measures ANOVA test was not possible due to the unequal number of data points by narration. At this point the results of the qualitative analysis is presented to direct the reflections. Their validity will need to be further tested with additional data.

6. Results

The results indicate that there are significant differences in the gesture frequency and duration by type of gesture but also by language. The following sections report the findings.

6.1 Gesture Frequency

The duration of each narration varied, the first English one (En1) was 3.1 minutes long, the second (En2) 1.84 minutes, similar to the Cantonese (Ct) narration 1.85 minutes. The number of words also varied (328 in the Cantonese narration, 446 for the first English one and 274 for the second English narration), as did the number of gestures (48 in Ct; 90 in En1 and 51 in En2). The frequency of gestures, measured as gestures per clause, decreases with proficiency, as expected (Table 1). Duration is given in seconds to facilitate reading although subsequent mentions are in milliseconds.

Table 1 Speech and gesture rates

P gestured throughout much of the three narrations (see Table 1), especially during the Ct narration (86% of the narration was observed with gestures). This is a higher proportion than in the English narrations where we observed gestures in 73% (En1) and 64% (En2) of the narration. We were surprised that the proportion would be higher in the less proficient English narration (En1) and went back to reanalyse the data, this time eliminating the duration of gestures on hold. During the En1 narration gestures were put on hold often, corresponding to speech disfluencies. Once gesture holds were excluded, the proportion of gesturing time in Cantonese was 80% and in both English narrations 62%. From this data, it would seem that P, although producing more individual gestures in less proficient narrations, actually spends more time gesturing in his mother tongue than in the L2.

We next analysed the duration (in milliseconds) of each gesture (including holds), confirming differences between the three narrations (Table 2). Mean gesture duration in Cantonese (M = 1987.69; SD = 977.77) was significantly longer than that in the English2 narration (M = 1400.33; SD = 714.21), and also longer than the gesture duration in the En1 narration (M = 1516.18; SD = 944.35). Gesture duration between the English narrations was not found to differ significantly (Table 2). Although the range of these mean durations is less than half a second, when observing P gesture the difference is noticeable even to an untrained eye (as confirmed by two independent helpers). The analysis confirmed that in Cantonese gestures were longer, slower and spanned a greater number of linguistic units, an average of 2.44 vs. 1.7 in the En1 narration and 2.05 in the En2 narration (if the gesture extended to part of a word this was also counted).

Table 2 Results from Paired Samples Wilcoxon signed-rank tests to check differences in gesture duration

Gestures co-occurred with pauses as well as with a variety of linguistic units. Overall, with proficiency we observed a trend to chunk more words under one gesture. In Cantonese, P's gestures often co-occurred with almost full clauses, while in the En1 narrations most of his gestures co-occurred with single word verbs or nouns. During the En2 narration we see the trend getting closer to that observed in the Cantonese narration (see Figure 1).

Fig. 1 Type of linguistic unit co-occurring with gestures

We confirmed that, overall, P produced more gestures in the less proficient narrations, calculated as the number of gestures per clause (see Table 1). There was an 8% decrease in gestures from the En1 narration (M = 0.375 gestures/clause, SD = 0.1) to the En2 (M = 0.345 gestures/clause, SD = 0.124) and a further 44% decrease from the En2 narration to the Cantonese one (M = 0.19 gestures/clause, SD = 0.085). As the proficiency increase had been relatively small (65% to 69.5%) the small reduction in gesture frequency between the En1 and En2 narrations is as expected.

6.2 Gesture Type

We sought to clarify whether P produced more gestures of a specific type. Gestures had been coded by their relationship to the content of the narration as either semantic (referential and deictic) or pragmatic (including beats and other gestures related to the discourse itself). The ratio semantic vs. pragmatic gesture is similar for the Cantonese and En1 narrations, close to 1, but in the En2 narration there are more pragmatic than semantic gestures, see Figure 2.

Fig. 2 Gestures per clause (by narration and gesture function)

We further analysed the semantic and pragmatic gestures dividing them into referential and deictic and beats and discursive gestures. There are similar decreases between referential and deictic gestures with proficiency. However,

the pattern is completely different with the pragmatic gestures. The frequency of beats increases slightly from the En1 to the En2 narrations, but this increase is more obvious in discursive gestures (see Figure 3). A comparison of the two English narrations (En1 vs. En2) and the En2 vs. Ct, indicates a similar decrease in referential gestures per clause (22% and 20% respectively). However, deictic gestures decrease 29% between the two English narrations but almost by half (49%) between the En2 and Cantonese narrations. Beats in the En1 vs. En2, increase by 2% and discursive gestures by 12%. Comparing the En2 narration with the Cantonese narration, we observe a 46% decrease in beats and a 61% decrease in discursive gestures. If the function of all gestures was similar, we would expect to see similar patterns in the changes of their frequencies. These results call for further studies separating gestures by their function but also observing how they change with the individual's increasing proficiency over time.

Fig. 3 Gestures per clause (by narration and gesture type)

We again turned to the duration (in milliseconds) of the gestures and compared these by type of gesture. Although for some categories the sample was small – affecting the reliability of the paired samples Wilcoxon signed-rank tests results – we believe these still provide some interesting insights:

- The average duration of deictic gestures is similar in all the narrations.
- There was a significant difference between the duration of referential gestures in the Cantonese (M = 2412.68, SD = 231.65) and En1 narrations (M = 1630.83, SD = 149.55): although with a large 95% CI [148.5, 1229], p = 0.004, Cohen's d = 0.856 (large effect) but not between the Cantonese and more proficient En2 narration (M = 1789.64, SD = 198.19) or between the two English narrations.
- There was a significant difference between the duration of discursive gestures in the Cantonese (M = 2193.42, SD = 299.35) and En1 (M = 1514.31, SD = 235.17): also with a broad 95% CI [500, 1919.5], p = 0.003, Cohen's d = 0.605 (medium effect); and also between the Cantonese and En2 narrations (M = 911.67, SD = 90.61) 95% CI [597, 1652.5], p = 0.001, Cohen's d = 1.64 (large effect). There was no significant difference between the two English narrations.

- There was a significant difference between the duration of beat gestures between the two English narrations En1 (M = 1215.526, SD = 645.148) and En2 (M = 1734.167, SD = 770.375): with a narrower 95% CI [-992.5, -208], p = 0.003, Cohen's d = 0.73 (medium effect).

The potential relevance of these findings is discussed below.

7. Discussion

Our analysis led to a number of observations that have been under-reported in L2 gesture research, to our knowledge. We describe these observations below and provide some possible explanations based on the following variables: proficiency, culture and differences in language type, such as syllabic duration and functional load as well as the syntactical structure. As our corpus is based on the narrations of just one individual, there is a great need for further studies with a larger number of participants, to validate or reject our observations. This discussion presents our results, comparing them with existing data on L2 gestures, and suggests possibilities for further research.

7.1 Overall frequency of gestures

P's frequency of gestures decreased with proficiency, confirming the results of most existing studies (Jacobsson, 2013; Gullberg & McCafferty, 2008, for a review; Lin, 2019; Stam, 2006; Zhao, 2006), Gregersen, Olivares-Cuhat and Storm (2009) are the exception – their study found more gestures in the L1 than the L2 of their participants. However, in terms of time spent gesturing, P actually gestured much more in his mother tongue than in the L2, accompanying over 86% of the narration with gestures (including 6% of hold time). Most gesture studies report on the number of individual gestures per unit of time or clause but future studies might also want to compare the total proportion of speech co-occurring with gestures. Lower proficiency speakers tend to produce many gestures of short duration, such as beats, rhythmic up-and-down movements of the hand. It is possible to see many of these gestures within one clause or in a minute, giving a high gesture frequency count. A more proficient speaker is more likely to produce longer gestures, giving a lower frequency count. However, a lower frequency count might not mean "less gesturing" because those gestures could be longer, so both speakers might be gesturing throughout the whole of their speech although producing a different number of gestures.

Reports of gestures per minute or second do not usually exclude pauses, which are likely to be longer and more common at lower levels of proficiency. Gestures are usually paused or interrupted with speech pauses (Graziano & Gullberg, 2018; Author, 2019) so the speech pause itself needs to be acknowledged when calculating overall speaking time. The text that follows refers to gesture frequencies per clause – although results were similar whether calculated as gestures per time or per clause – as it was more meaningful for our discussion.

7.2 Differences related to culture and language type

P was observed to produce more individual gestures per clause in the English narrations than in the Cantonese one. There is insufficient data to make any claims on the influence of culture on P's gestures. However, cultural factors are thought to play a role in the difference frequency of gestures in the L1 and the L2, as concluded by Iñigo-Mora and Álvarez-Benito (2012) who found Americans to gesture less than Spanish L1 speakers, or So (2010) who reported Mandarin speakers to gesture less than Americans (although neither study specifies the sub-cultural group of the American participants which is also relevant). Hou and So (2014) proposed that Chinese speakers might gesture less than American English speakers due to a Confucian-based culture which calls for body control. In studies comparing American and Chinese (Mandarin) caregivers to children, the latter produce more gestures when interacting with the children (Goldin-Meadow & Saltzman, 2000; So et al., 2014). Furthermore, four to six year old Chinese children were also found to gesture more than their American counterparts (So et al., 2010), suggesting culture might affect gesture production in later years. In another study carried out in the USA, So (2010) found that adult Mandarin-English bilinguals there produced a similar number of referential gestures when speaking in Mandarin and in English as English monolinguals. Both groups produced more referential gestures than Mandarin monolinguals, which So interpreted as related to a transfer from the American culture. In another study with bilinguals of English and various languages, frequency of gestures was found to be affected by cultural differences in story telling (Nicoladis et al., 2018). One important distinction is that these studies dealt with fully bilinguals and not language learners, but the overall implication is that culture does affect gesturing. Further research is needed in this area.

The gestures in the Cantonese narration were found to be significantly longer than those during either of the English narrations. Differences in gesture duration between L1 and L2 narrations might also be related to the functional load – the amount of information – carried by each syllable and the speech rate in each language. When spoken by native

speakers, Cantonese is a slower language than English, in terms of the information rate (Coupé, 2020), suggesting that gestures might be longer in Cantonese than in English. Indeed, the effect of syllable duration has been found to affect gestures in speakers of various languages (Pouw et al., 2020), although Cantonese was not one of the languages covered in the study. However, in P's case, the speech rate in English is affected by his level of proficiency, making it much slower than in Cantonese (also the case when measured as syllables per second). This should have meant slower gestures in En1 (with the slowest speech rate), which was not the case, suggesting there must be another factor at play.

In the less proficient En1 narration, P's gestures were the shortest, usually occurring with fewer than two linguistic units (mean of 1.7 units), and there were more of them. Some of these were related to disfluencies, where a gesture might start with a unit that was subsequently rephrased and the gesture abandoned; these were 13% of all gestures, vs. 2% in the En2 narration and none in the Cantonese. As illustrated in Figure 1, aside from those cases, gestures in the En1 narration spanned fewer linguistic units than in the En2 narrations (mean of 2.05), although the mean duration of most gestures was not significantly different. As the speech rate in the En2 narration was slightly (but not significantly) higher than in the En1 narration (see Table 1) this could explain why more speech units are covered by gesture. In the Cantonese narrations more linguistic units were covered per gesture (M = 2.45), and the speech rate was also faster than in the En2 narration.

Gesture duration might be affected by the structure of the language, Cantonese following a topic-comment structure (Matthews & Yip, 2011) that might lend itself to gestures co-occurring with the more elaborate comment part. The use of sentence end particles to clarify tense, mood or meaning could also lead to longer gestures that start with the verb, making the action salient, but continue until its tense or aspect is also highlighted by the gesture (this issue is currently being explored with a qualitative study including other Cantonese L1 speakers).

Differences in the structure of languages could also affect how the gesture is used (Slobin, 2006). With an action like "the cat rolls down the street", an English speaker would encode path and manner together into one gesture while a Chinese speaker would be more likely to just encode path (Kita & Özyürek, 2003). Therefore, if there is a need to make the manner salient as well this is likely to result in a second gesture, suggesting more referential gestures in Cantonese than in English. However, this is not what we observed. In the Cantonese narration, P produced just one

gesture per clause, referential or other. In the En1 narration there were 22 clauses with more than one gesture (out of 60, 37%) – six of those clauses co-occurred with more than one referential gesture – while in the En2 narration there were a similar percentage (38%, 14 clauses out of 37) but only two of those co-occurred with more than one referential gesture. None of the variables mentioned: culture, satellite or verb-framed language type, speech rate or syllable duration explain this observation. We believe it is more likely to be related to differences in proficiency.

7.3 Differences related to proficiency

We believe duration and gestures per clause differences to be related to how gesture-speech-idea are integrated and how this integration changes with proficiency. With proficiency (comparing the En1 and En2 narrations where speech rate was not significantly different), we observed that gestures span more linguistic units suggesting that the relationship speech-gesture is not limited to the linguistic unit but it is more likely to cover a chunk of speech. Multiword sequences are indicators of linguistic abilities (Arnon & Christiansen, 2017), and likely to develop with proficiency. We expect that whenever multiword sequences are processed as a chunk they will probably co-occur with just one gesture.

Modular based models of information processing propose that the manner in which the gesture (one module) extends across the speech (another module) may relate to the gesture-speech production mechanism (Levelt, 1993). If the gesture extends across the whole or most of the idea unit, then the gesture and the idea are likely to be conceptualized at the same time. However, if the gesture is limited to the extent of a single linguistic unit, then the origins of that gesture might relate to the linguistic production of the speech unit and the search for words to express the idea, rather than to the production of the idea itself. On the other hand, non-modular approaches bring together idea, speech and gesture as one (McNeill, 2016), supporting an integrated conceptualization.

We believe that the gesture-speech unity indicates how the L2 speaker is processing information, with gestures having a dual role: an external communicative one and an internal cognitive one. Speakers externalise mental models, the concept or idea, through idea-chunks, or clauses. Complex information might require a number of these chunks, which are designed to aid the interlocutor process them (Kita et al., 2017). Fluent speakers tie clauses with one key idea and the most salient element of that idea is often highlighted through a gesture (McNeill, 2005). However,

gestures also help speakers to conceptualize ideas (Church et al., 2004; Goldin-Meadow et al., 2009; Kita, 2000) including the spatial manipulation of the mental model (Kita et al., 2017) where gesture helps to consolidate the concept.

L2 speakers struggle with new concepts that might not exist in their L1 and also with the linguistic elements necessary to externalise their ideas. When L2 speakers experience conceptualization issues we would expect the gesture to reflect this as well, with pragmatic gestures that communicate lack of certainty (Graziano, 2020), same as a native speaker would, or with a pause in both speech and gesture (Graziano & Gullberg, 2018; Author, 2019). In P's narrations, we did not observe many failures of conceptualization; his narrations are true to the original input and detailed. As the input (the Tweety and Sylvester video) was given in a visual modality, the mental model formed would have been independent of the language to be subsequently tested – mental models are thought to be non-linguistic (Cutica & Bucciarelli, 2015).

Therefore, the observed speech-gesture differences between the three narrations should be partly related to the difference in proficiency levels. It is believed that differences in proficiency result in different degrees of speech processing automation (Kormos, 2014). This being the case, proficiency is also likely to affect gesture production and its integration with the speech (Author, 2019). Many studies with L1 speakers provide support for the integrated speech-gesture conceptualization theory (for a review see Kita et al., 2017). However, in P, and other L2 speakers, we observe behaviours that point to a modular gesture-speech conceptualization (ongoing study). We believe that there might be two different phases to the speech-gesture processing mechanism, an automated one that integrates idea-speech-gesture (as supported by McNeil, 2015), which is activated in the L1, and a modular phase where the gesture-idea are developed together but the linguistic element is added later, as observed in less proficient L2 speakers. This would support Hadar's model (2014), derived from the modular system, which suggest two stages of processing, one at the point of conceptualization and a follow on linguistic one. However, embodied cognition theories, closer to McNeills' gesture-speech-idea integration (2005), provide an alternative explanation. Barsalou and colleagues' "Language and Situated Simulation theory" (Barsalou et al., 2008, p. 253) refers to a linguistic system and a simulation one where a context-dependent model or simulation is created based on perception, action and introspection modal states. Modality-specific systems activate high level motoric and perceptual representations that interact with the

linguistic system, the simulations providing meaning to the linguistic forms (Barsalou et al., 2008). However, these two systems are not considered modular but dynamic, drawing on different brain networks to simulate different cognitive processes (Barsalou, 2008). If, in lower proficiency speakers, the cognitive process related to speaking is not as automated as in higher proficiency speakers, it is very likely that the networks activated might differ between the two and this might well be reflected in the gesture.

In the En1 narration, the less automated integration of the two systems or of the idea-speech-gesture results in a weaker gesture-speech link. This leads to additional gestures that provide redundant information (Negueruela & Lantolf, 2009). Proficient speakers are also likely to be fluent, with a good command of linguistic forms and thus able to externalise information that is effectively chunked and made salient through the gesture and prosody. This means each chunk or clause only needs one referential gesture. We believe that lower proficiency speakers, less able to control the linguistic chunking of information, are more likely to use more than one referential gesture per clause.

7.3.1 Differences by types of gestures

In P, semantic gestures decreased with proficiency, but gestures with a pragmatic function did not. Referential gestures were more frequent than other gestures in the Cantonese and En1 narrations, but in the En2 narration they were only the second most frequent, after discursive gestures. This narrative task lends itself to referential gestures, as it depicts concrete actions (rather than abstract thoughts). As there are a number of characters and repeated locations, we also expected a number of deictic gestures. Overall, we anticipated more semantic (referential and deictic) than pragmatic (beats and discursive) gestures in all three narrations. Instead, we observed an unpredicted increase in pragmatic gestures in the En2 narration, which we discuss below.

Studies comparing L2 and L1 speakers report a higher frequency of gestures in the L2 (Gullberg, 1999; Lin, 2019; Stam, 2006, Zhao, 2006). Zhao (2006) reported on different types of gestures observed in L1 and L2 interactions of L1 Chinese (Mandarin) speakers learning English. Participants were interacting with each other, which was reflected in an increase in pragmatic gestures between the L1 and the L2, seven-fold in beats and thirteen-fold in what Zhao categorizes as metaphorical attitudinal, while iconic gestures increased less than two-fold. Despite our task being a narration that did not require turn management or providing responses to an interlocutor, we also observed an increase

in pragmatic gestures in the L2, specifically in the more proficient narration, as also noted by Lin (2019) who compared two different groups of speakers with medium and high proficiency levels of English. Furthermore, all tasks were the same, so functionally, there was no need for additional pragmatic gestures in the En2 narration. Future studies should consider comparing narrations and interactions from the same speaker at different proficiency levels to confirm the role of the pragmatic gestures.

The frequency of referential and deictic gestures decreases with proficiency, although frequency of referential gestures in the En2 narration is closer to that in Cantonese than in the En1 narration. It has been suggested that lower proficiency speakers might produce lexical gestures (semantically related to content that cannot be externalised in speech) (Krauss et al., 2000) or additional referential gestures to elicit help from the interlocutor (Gullberg, 1998, 2006) or to establish temporal or anaphoric relationships (Gullberg, 1999, 2006). This might be the case at novice proficiency levels, but not with P. Instead, P opted to interrupt or put the gesture on hold while thinking for its lexical affiliate. P's proficiency was already high and his speech did not contain many vocabulary or grammar related difficulties or overall disfluencies, and these decreased in the second English narration. We did not, therefore, observe many gestures indicating a disfluency, lack of certainty or of vocabulary and these diminished with proficiency. There are also no instances where P asks the interlocutor for help with a gesture (Gullberg, 1999). We did observe more referential gestures in the lower proficiency narration, En1, often co-occurring with verbs (55% of all 29 referential gestures), this frequency seeming to converge in the En2 and Cantonese narrations (29% out of 14 and 26% out of 19 referential gestures respectively) where the linguistic system, including word chunking, is thought to be more automated. We did not carry out a semantic comparison of the speech and the gesture so we cannot confirm whether the gestures are adding supplementary information but we expect there is more gesture-speech redundancy at the lower proficiency level in an effort to ensure that the information is being conveyed successfully.

Gullberg (2003, 2006) had reported this redundancy with deictic gestures. We also observed a higher frequency of deictic gestures in the En1 narration, decreasing with proficiency. Deictic gestures act as visual referents, connecting parts of the discourse by indicating their position in space, these gestures usually co-occur with anaphoric references. Lower proficiency speakers who have not mastered the use of pronouns and other anaphoric references will often repeat noun phrases (Gullberg, 2006), with their corresponding gestures or use them to establish temporal relations

(Gullberg, 1999). As proficiency develops, repeated noun phrases are eliminated or replaced with anaphoric references, which is reflected in fewer gestures. Some of the deictic gestures co-occurred with the abstract location of the characters. These were often used in the English narrations instead of specifying the location in the speech.

A recent study by Debreslioska and Gullberg (2020) notes that speakers use more gestures with inferable referents, those that are new to the interlocutor but known from the previous discourse. The En1 narration came after the Cantonese task, where the first part of the video was watched and retold (existing discourse), therefore it is possible that our task set-up had the effect of increasing the frequency of the En1 referential gestures. Future studies should also account for this variable.

In the analysis, pragmatic gestures were categorized as beats or discursive as we expected to see a difference between the two. Lower proficiency speakers have been observed to use beats to place in space the various parts of a sentence (McCafferty, 2006), parsing its structure. They also often beat rhythmically with each syllable, resulting in a string of up and down movements, often as speech is corrected (Gullberg, 1998). With P we confirmed there were more beats in the English narrations than in the Cantonese one but a similar frequency of beats in both English narrations. The duration of beats in the En1 narration was shorter than in En2 narrations, suggesting that they might play a different function than the beats in the L2. Based on the work by McCafferty (2006) it is possible that many of the En1 beats might be related to syllable separation rather than to marking discursive stress. The high frequency of beat gestures in the En2 narration was unanticipated, as they were not used as a parsing aid or to correct speech. One possible explanation is that they were marking the prosody (Im & Baumann, 2020). The proficiency evaluators had indicated that intonation in particular had improved in P, perhaps as P became more confident in his syntax and lexicon he began to pay more attention to prosody and used beats to mark it. Future research should also look for changes in the functions of beats with proficiency.

Gestures with pragmatic functions manage the interaction, the flow of information or indicate inference, causality or other types of links between utterances. Discourse gestures have similar functions to those of speech discourse markers: connecting utterances (and, then); indicating interaction (hm?); providing logical connections between clauses (because, but); indicating how the speaker feels about the utterance (I think), among others (see González, 2004, for a review of L1 and L2 discourse markers in oral discourse). Discourse gestures are often recurrent and can

be observed across speakers and contexts with a similar function and content (Ladewig, 2014) in various languages (Author, 2020). In the En1 narration 20% of the 26 discourse gestures (5 cases) co-occur with discourse markers, in the En2 narration this goes up to 30% (5 cases out of 18, excluding two discourse markers with non-discursive gestures). In the Cantonese narration, half of the 12 discursive gestures observed co-occur with discourse markers (4 cases) or parts of clauses containing them (2 cases).

The frequency and duration of discursive gestures also differed between narrations. We believe that as proficiency develops, clauses and the cognitive relations between them (causality, consequence, finality, inference), become more complex. Gestures will be used to indicate these relations while the linguistic skills are being developed. The synchrony gesture-discourse marker increases with proficiency as the speaker acquires the various pragmatic functions of discourse markers, which are just indicated by gestures in lower proficiency speakers. P's proficiency evaluators also found improvements in P's fluency, coherence and parallelism, all factors closely related to the overall flow of the narration and noted by an increase in the pragmatic functions of discourse markers. The pragmatic functions of discourse markers in speech are quite complex (Borreguero Zuloaga, 2015) and their link to gestures deserves further research.

8. Summary

Our results confirm those of previous studies with speakers of different proficiency levels (Lin, 2019; Zhou, 2006) that point to differences in how gestures are produced at different points in the proficiency continuum and which note that these changes might not be linear. Our results with the same speaker at different proficiency levels (rather than different speakers at different levels) confirm that referential and deictic gestures might be the first to resemble L1 referential patterns in terms of frequency and gesture duration, in this speaker. On the other hand, discursive gestures seem to provide a better indication of the proficiency level of the speaker (at least in this case).

Based on our observations of P, we suggest that when speakers develop mental models the non-linguistic concepts are created regardless of the language to be used to express those concepts. When externalisation takes place, the 'one idea-one linguistic chunk' integration improves with proficiency. In the lower proficiency levels, where this integration occurs at unit rather than chunk level, speakers are more likely to employ individual gestures to highlight

each linguistic unit, rather than a specific unit within a chunk. Therefore, there are more referential gestures which are expected to be superfluous, not adding information. This would indicate that the two modules, gesture and speech, are not fully integrated as they are externalising the same information. As proficiency develops the frequency of referential gestures decreases and, our expectation is that they will also be more complementary, making the speech-gesture more efficient, in terms of delivering information.

Meanwhile, with proficiency, there is also an improvement in how the overall narration is structured. Lower proficiency speakers use shorter sentences and a limited range of connectors. As speakers become more proficient, from an A2 level (CEFR), they attempt to provide a more complex structure to their utterances, with more discourse markers to connect their ideas. This results in an increase of discourse related gestures, which are not always produced with a related lexical affiliate, as the speaker tries to add pragmatic meaning to their utterances. We suggest that it is only in the higher proficiency levels, C1 and above – but sometimes also in B2 speakers who have acquired the language in an immersion context – and native speech, that we will observe the full integration of discourse gestures and lexical affiliates, like discourse markers, which provide the interlocutor with the impression of 'fluent speech'.

9. Conclusions

The observations from the L1 and L2 narrations of this Cantonese speaker leave us with no doubt as to the need for further longitudinal research into the gestures of L2 learners to better understand how they develop in individuals with increasing proficiency. Our observations clearly suggest that there are a number of variables that need to be taken into consideration when comparing gestures in different languages, not just the proficiency levels. One of these is the language type, including: the amount of information carried by each syllable, the duration of each syllable, the structure of the language in terms of how and what information is presented. Cultural factors might also play a role; in relation to this study it has been observed that Chinese English speakers gesture less than their American counterparts (So, 2010; So et al., 2010) albeit our speaker was a Cantonese and not a Mandarin L1 speaker and based in Hong Kong, not the USA. Although not tested in this study, it is also likely that the task being performed (narrations vs. interactions) will affect gesture rate and function. Other factors that are have been noted to play a role in gesture production include individual cognitive abilities relating to spatial cognition and verbal and non-verbal memory (McKern et al., 2021). We would also like to suggest that creativity (G. Stam, personal communication, October 2021)

and other personality aspects, such as confidence or extraversion might also influence gesture production. Longitudinal studies tracking changes in proficiency in specific individuals will allow for the control of language type and culture variables as well as idiosyncratic differences. Based on our observations, and those of previous studies, there are at least two main areas of future research:

1. In higher proficiency speakers the frequency of referential gestures in the L2 seems to converge with frequencies observed in the L1. This might be related to a more effective use of gesture-word chunking and should be tested against changes in word chunking and gesture-speech redundancy.

2. There is an increase in gestures with pragmatic functions (beats and discursive gestures) as proficiency increases (at least from high intermediate to high proficiencies). More data is required, covering different proficiency levels. We believe beats might increase to reflect prosodic control while discursive gestures increase as they accompany more complex narrations. However, it is not clear whether they peak and then decrease in higher proficiency speakers or continue to increase.

The practical implications of understanding the gestures of language learners extends to the language classroom and to language proficiency evaluation. The rubrics for evaluating proficiency according to the Common European Framework of Reference for Languages only account for non-verbal cues when assessing the highest level, C2. However, gestures (and other non-verbal cues such as gaze, body or head gestures) are present in most language learners, at all levels. These ought to be taken into account when assessing all oral proficiency levels, as they contribute to the communicative success of any exchange, providing coherence and cohesion as well as very important pragmatic information (Author, 2020). Gestures also need to be further studied to understand their cultural role in the L1 and the L2, how they are transferred and perceived by interlocutors who might or might not share the same languages as the speaker. Finally, and the topic of a much larger longitudinal project currently ongoing, we need to understand gesture function at each proficiency level.

Further research into gestures in teaching and learning is needed to understand not only how teachers' gestures influence learning (if at all) but also how, and when, learners develop a full speech-gesture-conceptualization integration. The inclusion of gestures as an additional resource in vocabulary and grammar teaching is thought to aid

learning (Cutica & Bucciarelli, 2015; Macedonia, 2019; Matsumoto & Dobs, 2017). Producing relevant gestures as part of the learning process is also reported to be beneficial (Cutica et al., 2014; Engelkamp & Jahn, 2003). However, how these gestures are integrated as part of the conceptualization of the content being learnt is still not clear. We believe that understanding the link speech-gesture-conceptualization at difference proficiency levels will allow language pedagogues to improve not only their teaching but also the evaluation of proficiency levels.

References

Author, 2019

Author, 2020

Andric M, Small S.L. (2012). Gesture's Neural Language. *Frontiers in Psychology*, *3*, 99. doi:10.3389/fpsyg.2012.00099.

Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1–L2 differences. *Topics in Cognitive Science*, *9*(3), 621-636.

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing In M. De Vega, A. Glenberg & A. Graesser (Eds.), *Symbols and embodiment: Debates on meaning and cognition* (pp. 245-283). Oxford University Press.

Bernstein, J., Van Moere, A., & Cheng, J. (2010). Validating automated speaking tests. *Language Testing*, *27*(3), 355-377.

Billot-Vasquez, K., Lian, Z., Hirata, Y., & Kelly, S. D. (2020). Emblem Gestures Improve Perception and Evaluation of Non-native Speech. *Frontiers in psychology*, 11, 2351.

Boersma, P., & Weenink, D. (2019). *Praat: Doing Phonetics by Computer*. Version 6.0.54 [Computer program]. http://www.praat.org/. Accessed 6 November 2020.

Borreguero Zuloaga, M. (2015). A vueltas con los marcadores del discurso: de nuevo sobre su delimitación y sus funciones. In A. Ferrari, L. Lala & R. Stojmenova (Eds.), *Testualità. Fondamenti, unità, relazioni/Textualité. Fondements, unités, relations/Textualidad. Fundamentos, unidades, relaciones* (pp. 151-170). Cesati, Firenze.

Brown, A., & Gullberg, M. (2008). Bidirectional crosslinguistic influence in L1-L2 encoding of manner in speech and gesture: A Study of Japanese Speakers of English. *Studies in Second Language Acquisition*, *30*, 225–251. http://doi.org/10.1017/S0272263108080327.

Chui, K. (2005). Temporal patterning of speech and iconic gestures in conversational discourse. *Journal of Pragmatics*, *37*(6), 871-887. https://doi.org/10.1016/j.pragma.2004.10.016.

Church, R. B., Ayman-Nolley, S., & Mahootian, S. (2004). The role of gesture in bilingual education: Does gesture enhance learning?. *International Journal of Bilingual Education and Bilingualism*, *7*(4), 303-319.

Cobb, T. (2002). *Web Vocabprofile an adaptation of Heatley, Nation & Coxhead's.* http://www.lextutor.ca/vp/. Accessed 6 November 2020.

Coupé, C. (2020). Carrying information with speech. Talk presented at the *Research Centre for Language, Cognition, and Neuroscience Research Seminar* series (10 Sept. 2020). The Hong Kong Polytechnic University.

Cutica, I., & Bucciarelli, M. (2015). Non-determinism in the Uptake of Gestural Information. *Journal of Nonverbal Behavior*, *39*(4), 289-315.

Cutica, I., Ianì, F., & Bucciarelli, M. (2014). Learning from text benefits from enactment. *Memory & Cognition*, *42*(7), 1026-1037.

Debreslioska, S., & Gullberg, M. (2020). What's new? Gestures accompany inferable rather than brand-new referents in discourse. *Frontiers in Psychology*, *11*.

Engelkamp, J., & Jahn, P. (2003). Lexical, conceptual and motor information in memory for action phrases: A multi-system account. *Acta Psychologica*, *113*(2), 147-165.

Freleng, F. (director) (1950). *Canary row* [Animated Film]. New York: Time Warner.

Fung, H. S. H., & Mok, P. P. K. (nd.). *A Preliminary study of the temporal relationship between prosody and gesture in Hong Kong Cantonese*. Retrieved from: Academia.edu.

Gluhareva, D., & Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research, 21*(5), 609-631. http://dx.doi.org/10.1177/1362168816651463.

Goldin-Meadow, S., & Saltzman, J. (2000). The cultural bounds of maternal accommodation: How Chinese and American mothers communicate with deaf and hearing children. *Psychological Science, 11*(4), 307-314.

Goldin-Meadow, S., Wagner-Cook, S., & Mitchell, Z. A. (2009). Gesturing gives children new ideas about math. *Psychological Science, 20*(3), 267-272.

González, M. (2004). *Pragmatic markers in oral narrative: The case of English and Catalan* (Vol. 122). Netherlands: John Benjamins Publishing.

Graziano, M. (2020). The interplay between speech and gesture in children's language development. Paper presented at the *7th Gesture and Speech in Interaction Conference*. KTH Speech, Music & Hearing and Språkbanken Tal, Stockholm, Sweden (September 2020).

Graziano, M., & Gullberg, M. (2018). When speech stops, gesture stops: Evidence from developmental and crosslinguistic comparisons. *Frontiers in Psychology, 9*, 879.

Gregersen, T., Olivares- Cuhat, G., & Storm, J. (2009). An Examination of L1 and L2 Gesture Use: What Role Does Proficiency Play? *The Modern Language Journal, 93*(2), 195–208. https://doi.org/10.1111/j.1540-4781.2009.00856.x.

Grice, H. P. (1975). Logic and conversation. In Cole, P., & Morgan, J. (Eds.). *Speech Acts* (pp. 41-58). Netherlands: Brill.

Gullberg, M. (1998). Gesture as a communication strategy in second language discourse: A study of learners of French and Swedish. *Travaux de l'Institut de Linguistique de Lund*, 35. Lund, Sweden: Lund University.

Gullberg, M. (1999). Gestures in spatial descriptions. *Travaux de l'Institut de Linguistique de Lund*, 87-98. Lund, Sweden: Lund University.

Gullberg, M. (2006). Handling discourse: Gestures, reference tracking, and communication strategies in early L2. *Language Learning, 56*(1), 155-196.

Gullberg, M. (2009). Reconstructing verb meaning in a second language: How English speakers of L2 Dutch talk and gesture about placement. *Annual Review of Cognitive Linguistics, 7*(1), 221-244.

Gullberg, M. (2011). Thinking, Speaking and Gesturing. In Pavlenko, A. (Ed.). *Thinking and speaking in two languages* (pp. 143-169). Bristol: Multilingual Matters.

Gullberg, M., & McCafferty, S. G. (2008). Introduction to Gesture and SLA: Toward an Integrated Approach. *Studies in Second Language Acquisition, 30*(2), 133–146.

Hadar, U. (2014). Coverbal gesture: Between communication and speech production. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body-Language-Communication* (Vol. 2, pp. 804-821). Berlin: de Gruyter.

Heatley, A., Nation, I.S.P., & Coxhead, A. (2002). *RANGE and FREQUENCY programs*. http://www.victoria.ac.nz/lals/staff/paul-nation.aspx . Accessed 6 November 2020.

Hou, S., & So, W. C. (2014). Gestures in China: Universal and culturally specific characteristics. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body-Language-Communication* (Vol. 1, pp. 1233-1239). Berlin: de Gruyter.

Igualada, A., Esteve-Gibert, N., & Prieto, P. (2017). Beat gestures improve word recall in 3-to 5-year-old children. *Journal of Experimental Child Psychology, 156*, 99-112. http://dx.doi.org/10.1016/j.jecp.2016.11.017.

Im, S., & Baumann, S. (2020). Probabilistic relation between co-speech gestures, pitch accents and information status. *Proceedings of the Linguistic Society of America, 5*(1), 685-697.

Iñigo-Mora, I., & Álvarez-Benito, G. (2012). Gesture and speech in topic organization in English as L1 and L2. *International Journal of Humanities and Social Science, 2*(6), 70-102.

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct?. *Applied linguistics, 29*(1), 24-49.

Jacobsson, G. A comparative study of speaker's gestures in L1 versus L2 speech in students studying Swedish at Sfi/Komvux. Unpublished report. Lund University. https://www.researchgate.net/publication/260210111. Accessed 6 November 2020.

JASP Team (2020). *JASP* (Version 0.14)[Computer software]. Accessed 6 November 2020.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture: Window into thought and action* (pp. 162-185). New York: Cambridge University Press.

Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: A review. *Language and cognitive processes, 24*(2), 145-167.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48*, 16-32.

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological review, 124*(3), 245.

Kok, K., Bergmann, K., Cienki, A., & Kopp, S. (2016). Mapping out the multifunctionality of speakers' gestures. *Gesture, 15*(1), 37-59.

Kormos, J. (2006). *Speech production and second language acquisition*. New Jersey: Lawrence Erlabaum Associates.

Kormos, J. (2014). *Speech production and second language acquisition.* New York: Routledge.

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: a process. In D. McNeill (Ed.), *Language and gesture: Window into thought and action* (pp. 261-283). New York: Cambridge University Press.

Kushch, O., Igualada, A., & Prieto, P. (2018). Prominence in speech and gesture favour second language novel word learning. Language, *Cognition and Neuroscience*, *33*(8), 992-1004.

Kwan, C. Y., Kwan, C., & Kong, P. (2012). *A systematic quantification of gesture employment in normal Cantonese speakers: A multi-directional analysis of forms and functions*. Unpublished dissertation, The University of Hong Kong, Hong Kong SAR, China. http://hub.hku.hk/handle/10722/237905. Accessed 6 November 2020.

Ladewig, S. H. (2014). Recurrent gestures. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body-Language-Communication* (Vol. 2, pp. 1558-1575). Berlin: de Gruyter.

Levelt, W. J. (1993). *Speaking: From intention to articulation* (Vol. 1). MA.: MIT press.

Lin, Y. L. (2019). Speech-accompanying gestures in L1 and L2 conversational interaction by speakers of different proficiency levels. *International Review of Applied Linguistics in Language Teaching*, 1 (ahead-of-print).

Lin, Y. L. (2020). A helping hand for thinking and speaking: Effects of gestures and task planning on second language narrative discourse. *System*, 102243.

Neu, J. (1990). Assessing the role of nonverbal communication in the acquisition of communicative competence in L2. In R. C. Scarcella, E. S. Andersen & S. D. Krashen (Eds.), *Developing Communicative Competence in a Second Language* (pp. 121–138). New York, NY: Newbury House.

Macedonia, M. (2019). Embodied Learning: Why at school the mind needs the body. *Frontiers in Psychology*, *10*, 2098.

Matsumoto, Y., & Dobs, A. M. (2017). Pedagogical gestures as interactional resources for teaching and learning tense and aspect in the ESL grammar classroom. *Language Learning, 67*(1), 7-42.

Matthews, S., & Yip, V. (2011). *Cantonese: A comprehensive grammar* (Second ed., Routledge comprehensive grammars). Abingdon, Oxon: Routledge.

McCafferty, S. G. (2004). Space for cognition: Gesture and second language learning. *International Journal of Applied Linguistics, 14*(1), 148-165.

McCafferty, S. G. (2006). Gesture and the materialization of second language prosody. International *Review of Applied Linguistics in Language Teaching, 44*(2), 197-209.

McCafferty, S. G. (2009). Material foundations for second Language acquisition. In S. McCafferty & G. Stam (Eds.), *Second Language Acquisition and classroom research* (pp. 47–65). New York: Routledge.

McCafferty, S., & Gullberg, M. (2008). Gesture and SLA-Toward an integrated approach. *Studies in Second Language Acquisition, 30*(2), 133-276.

McKern, N., Dargue, N., Sweller, N., Sekine, K., & Austin, E. (2021). Lending a Hand to Storytelling: Gesture's Effects on Narrative Comprehension Moderated by Task Difficulty and Cognitive Ability. *Quarterly Journal of Experimental Psychology, 74*(10), 1791-1805.

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

McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.

McNeill, D. (2016). *Why we gesture*. New York, NY: Cambridge University Press.

Morrel-Samuels, P., & Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(3), 615.

Müller, C. (1998). *Redebegleitende Gesten. Kulturgeschichte - Theorie - Sprachvergleich*. Berlin: Berlin Verlag Arno Spitz GmbH.

Müller, C. (2004). Forms and uses of the Palm Up Open Hand: a case of a gesture family. In: Müller, C., Posner, R. (Eds.), *The Semantics and Pragmatics of Everyday Gestures* (pp. 233-256). Weidler, Berlin.

Negueruela, E., & Lantolf, J. P. (2009). The dialectics of gesture in the construction of meaning in second language oral narratives. In S. G. McCafferty & G. Stam (Eds.), *Gesture: Second language acquisition and classroom research* (pp. 88-106). Oxfordshire, UK: Routledge Taylor & Francis Group.

Nicoladis, E., Nagpal, J., Marentette, P., & Hauer, B. (2018). Gesture frequency is linked to story-telling style: Evidence from bilinguals. *Language and Cognition, 10*(4), 641-664. doi:10.1017/langcog.2018.25.

Nicoladis, E., Pika, S., & Marentette, P. (2009). Do French–English bilingual children gesture more than monolingual children?. *Journal of Psycholinguistic Research, 38*(6), 573-585.

Nicoladis, E., Pika, S., Yin, H. U. I., & Marentette, P. (2007). Gesture use in story recall by Chinese-English bilinguals. *Applied Psycholinguistics, 28*(4), 721.

Pika, S., Nicoladis, E., & Marentette, P. F. (2006). A cross-cultural study on the use of gestures: Evidence for cross-linguistic transfer?. *Bilingualism, 9*(3), 319.

Pouw, W., Jaramillo, J., Ozyurek, A. & Dixon, J. (2020). Quasi-rhythmic features of hand gestures show unique modulation within languages: Evidence from bilingual speakers. Paper presented at the *7th Gesture and Speech in Interaction Conference*. KTH Speech, Music & Hearing and Språkbanken Tal, Stockholm, Sweden (September 2020).

Rohrer, P. L., Delais-Roussarie, E., & Prieto, P. (2020). Beat gestures for comprehension and recall: Differential effects of language learners and native listeners. *Frontiers in Psychology,* 11. doi:10.338/fpsyg.2020.575929.

Saddour, I. (2017). A multimodal approach to investigating temporality expression in L2: What does gesture analysis reveal? *International Review of Applied Linguistics in Language Teaching, 55*(3), 283.

Scott, M. (2004). *WordSmith Tools* version 4, Oxford: Oxford University Press.

Shattuck-Hufnagel, S., & Prieto, P. (2019). Dimensionalizing co-speech gestures. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (Eds.), Proceedings of the *International Congress of Phonetic Sciences Melbourne* (pp. 1490-1494). Canberra: Australia Speech Science and Technology Association Inc.

Sherman, J., & Nicoladis, E. (2004). Gestures by advanced Spanish-English second-language learners. *Gesture, 4*(2), 143-156.

Slobin, D. I. (2006). Explorations in linguistic typology, discourse, and cognition. In M. Hickmann & S. Robert (Eds.), *Space in languages: Linguistic systems and cognitive categories* (Vol. 66, pp. 59-80). Amsterdam: John Benjamins.

Sloetjes, H. (2017). *ELAN* [Computer program]. Version 5.0.0. https://archive.mpi.nl/tla/elan. Accessed 6 November 2020.

So, W. C., Kita, S., & Goldin-Meadow, S. (2013). When do speakers use gestures to specify who does what to whom? The role of language proficiency and type of gestures in narratives. *Journal of Psycholinguistic Research, 42*(6), 581-594.

So, W. C. (2010). Cross-linguistic transfer in gesture in Chinese-English bilinguals. *Language and Cognitive Processes*, *25*(10), 1335-1353.

So, W. C., Demir, E, & Goldin-Meadow, S. (2010) When speech is ambiguous gesture steps in: Sensitivity to discourse-pragmatic principles in early childhood. *Applied Psycholinguistics, 31(*1), 209-224.

So, W. C., Lim, J.Y., & Tan, S.H. (2014). Sensitivity to information status in discourse: Gesture precedes speech in unbalanced bilinguals. *Applied Psycholinguistics, 35*, 71-95.

Stam, G. (2006). Thinking for speaking about motion: L1 and L2 speech and gesture. *IRAL International Review of Applied Linguistics in Language Teaching, 44*(2), 145–171.

Stam, G. (2015). Changes in Thinking for Speaking: a longitudinal case study. *The Modern Language Journal, 15*, 83-99.

Talmy, L. (2000). *Toward a cognitive semantics* (Vol. 2). M.A.: MIT press.

Zhao, J. (2006). The communicative functions of gestures in L2 Speech. *Journal of Second Language Acquisition and Teaching, 13*, 1-17.

Fig. 1 Type of linguistic unit co-occurring with gestures

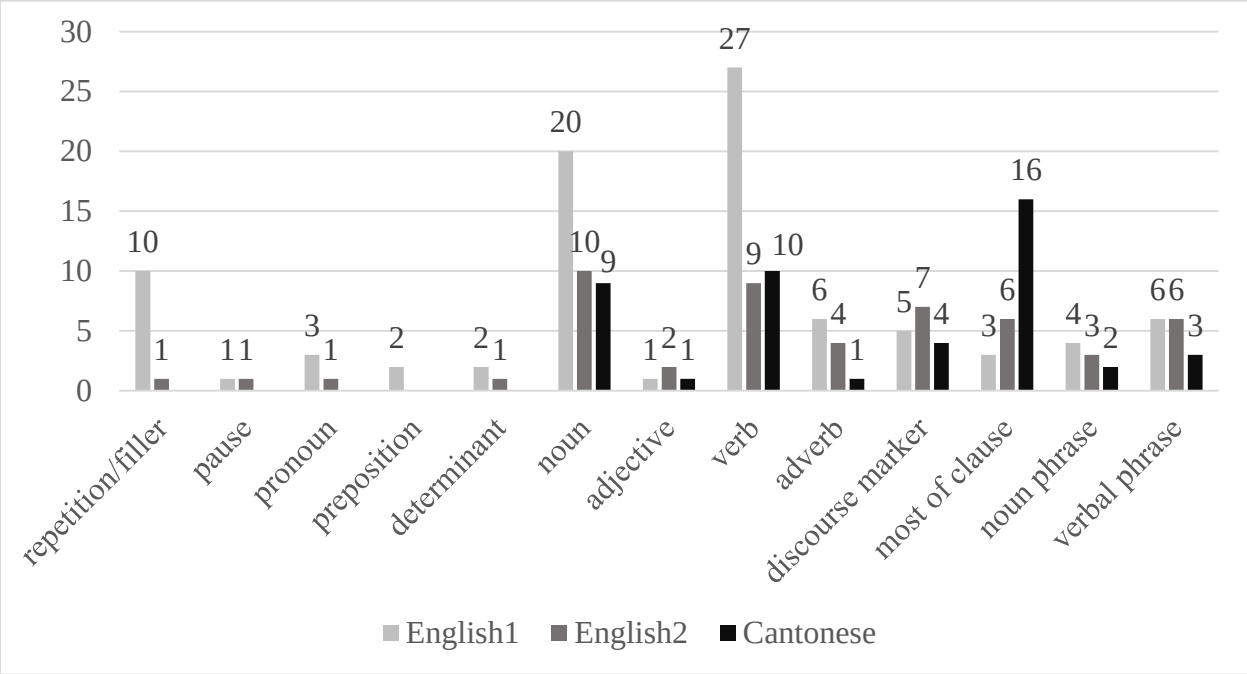

Fig. 2 Gestures per clause (by narration and gesture function)

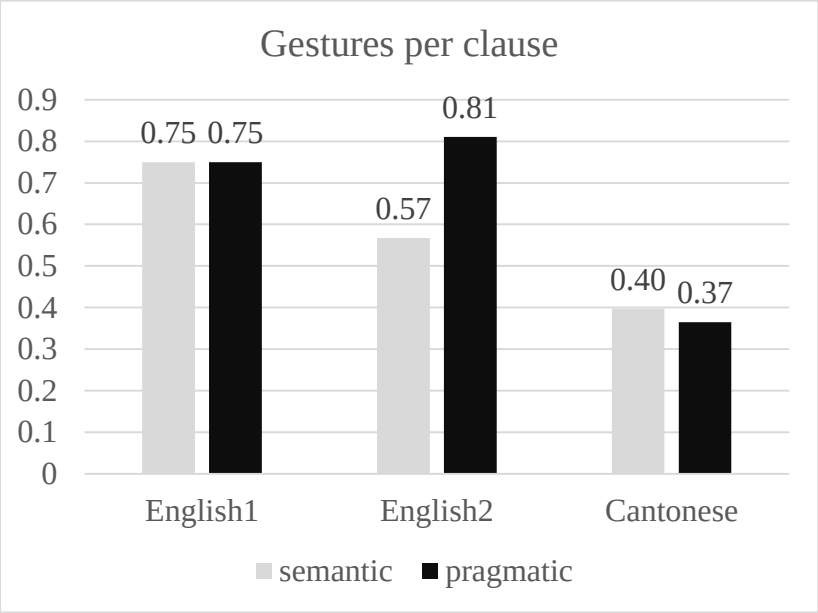

Fig. 3 Gestures per clause (by narration and gesture type)

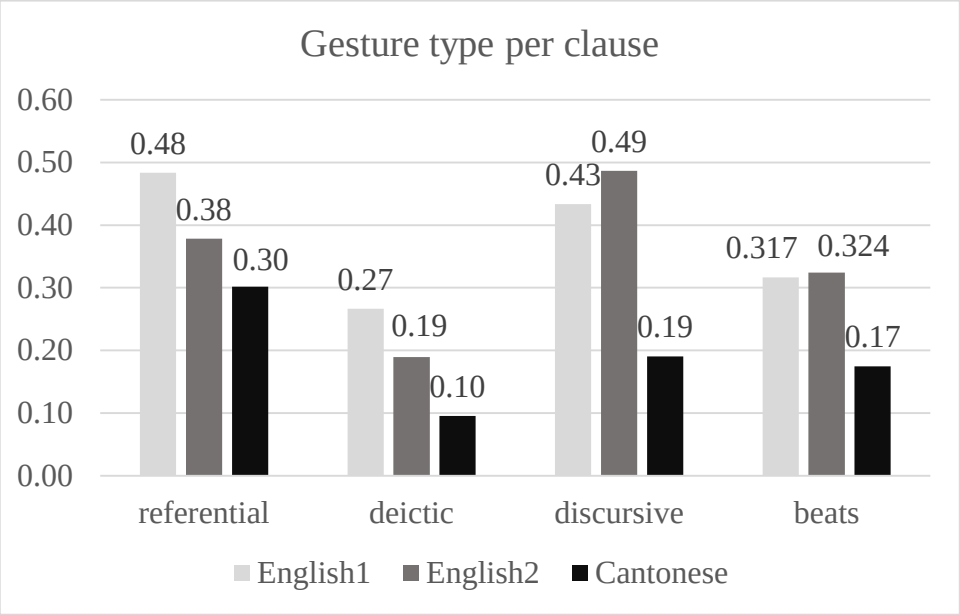

Table 1 Speech and gesture rates

	Speech rate (words/ second)	Total gestures	Total clauses	Gestures/ clause	Duration of narration (seconds)	Time spent gesturing (seconds)	% time gesturing (including holds)	% time gesturing (excluding holds)
En1	2.392	90	60	1.5	186.438	136.456	73%	62%
En2	2.472	51	37	1.378	110.853	71.417	64%	62%
Ct	2.955	48	63	0.762	111.000	95.409	86%	80%

Table 2 Results from Wilcoxon signed-rank test to check differences in gesture per clause

Dependent Pairs	W	p	95% CI for Mean Difference	Cohen's d
Cantonese (N = 48) – English 1 (N = 90)	807	0.024*	[41.5, 687.5]	0.49 (medium)
Cantonese (N = 48) – English 2 (N= 51)	860	0.005*	[165, 869.5]	0.686 (medium)
English 1 (N = 90) – English 2 (N= 51)	774.5	0.298	[-154.5, 501.5]	0.138 (low)

Notes: * marks significant differences at alpha = 0.05