

# Co-occurrence of speech and gestures: A multimodal corpus linguistic approach to intercultural interaction<sup>☆</sup>



Yen-Liang Lin

Department of English, National Taipei University of Technology, 1, Sec. 3, Zhongxiao E. Rd, Taipei, Taiwan

Received 2 September 2016; received in revised form 19 June 2017; accepted 25 June 2017

## Abstract

This study investigates face-to-face interaction among Taiwanese, Indonesian, and Indian speakers utilizing a multimodal corpus linguistics approach to examine semantic categories of speech that most frequently co-occur with gestures, and whether the gesture-speech relationship is to a certain extent influenced by language/culture backgrounds or English proficiency levels of a speaker. The analysis of the semantic categories of the co-gesture speech demonstrates that speech most commonly co-occurs with gestures in the categories of moving, coming and going, general objects, numbers, location and direction, and time. The findings demonstrate similar preferences of gesture-speech production by speakers despite different cultural and linguistic backgrounds. The gesture-speech relationship was shown to fall into six discrete categories: *reinforcing*, *integrating*, *supplementary*, *complementary*, *contradictory*, and *others*. While results show that the gesture-speech relationship is not significantly influenced by different language backgrounds of a speaker, speakers at a high proficiency level tended to use significantly more gestures that serve *reinforcing* and *integrating* functions, whereas less proficient speakers produced more gestures as *complements* and *other* gestures that have no obvious relationship to the conceptual content of their accompanying speech.

© 2017 Elsevier B.V. All rights reserved.

**Keywords:** Speech-gesture relationship; Multi-modal corpora; Semantic domains; Corpus linguistic approach; Co-gesture speech

## 1. Introduction

Speaker gestures are semantically co-expressive with speech and serve different pragmatic functions to accompanying speech (Holler and Wilkin, 2011; Kendon, 2004; McNeill, 1992, 2005, 2016; Stam and McCafferty, 2008). When one looks at the gesture-speech relationship, it can be seen that “gesture orchestrates speech” in many different ways (McNeill, 2016, p. 4). For example, gestures encode both propositional and imagistic information and may further reinforce the information in the accompanying speech, in that they convey similar semantic content as speech (McNeil et al., 2000). Gestures may also provide important information or significance to the accompanying speech and add clarity to discourse (Colletta et al., 2015). Moreover, gestures can be employed to facilitate lexical retrieval and retain a turn in conversations (Stam and McCafferty, 2008) and also assist in verbalizing semantic content (Alibali et al., 2007). In this regard, as noted by Morsella and Krauss (2004), gestures facilitate speakers in coming up with the words they intend to say by sustaining the activation of a target word's semantic features long enough for the process of word production to take place. The tight link between speech and gesture can be explained by two gesture-speech characteristics: semantic coherence (combining gesture with meaningful and related speech) and temporal synchrony (producing gesture in

---

<sup>☆</sup> Data collection and analysis for this study was financed by the Ministry of Science and Technology, Taiwan (Project 104-2410-H-027-001).  
E-mail address: [ericlin@ntut.edu.tw](mailto:ericlin@ntut.edu.tw).

synchrony with speech) (Butcher and Goldin-Meadow, 2000, p. 150). Semantic coherence refers to co-expressivity of gesture and speech, in which the modalities are semantically coordinated and mostly work together to convey a single thought unit (McNeill, 1992, 2016; Kendon, 2004). This supports McNeill's (1992, 2016) claim that gesture and speech share similar cognitive representation and are therefore a unified system. Temporal synchrony refers to the fact that the two modalities occur at or almost at the same time. McNeill (1992) found that up to 90% of gestures are aligned with speech co-temporally. Even when communicators stutter, gesture and speech are still used synchronously (Mayberry and Jaques, 2000). It appears that there is strong evidence supporting the idea that there exists a concomitant relationship between the two modalities in terms of semantic and temporal coordination. While recent studies have provided anecdotal and experimental evidence, looking at different semantic features of speech accompanied with gestures, such as path and motion events (e.g., Arslangul, 2015), locative information (e.g., Tutton, 2011), reference and deictic expressions (So et al., 2013), and spatial properties (e.g., Johansson and Zlatev, 2013) of gesture and speech, few, if any, have systematically examined the semantic components of gesture-speech co-occurrence using a corpus linguistic approach.

Although current approaches in corpus linguistics have revolutionized how we look at spoken discourse, most corpora are limited in the sense that they only have the provision for presenting data in a single format and provide few opportunities for exploring non-verbal, gestural features of discourse, which are important aspects of understanding intercultural face-to-face interaction (Adolphs and Carter, 2013; Knight, 2011). Consequently, in such a format the “data” may lose some of the characteristics (particularly non-verbal) that existed as part of the spoken discourse itself, characteristics which can play an integral part in determining the function and meaning of linguistic units in conversation. In support of this, Adolphs and Carter (2013) argue that “spoken interaction is essentially multimodal in nature, featuring a careful interplay between textual, prosodic, gestural and environmental elements in the construction of meaning” (p. 4). It is, therefore, necessary to propose a new generation of corpora, accommodating a more multimodal perspective of discourse. Moreover, non-verbal behavior has been viewed as being part of strategic competence in second or foreign language communication, as explored by Canale (1983) who suggested that strategic competence “is composed of mastery of verbal and non-verbal communication strategies in order to compensate for breakdowns in communication due to insufficient competence and to enhance the effectiveness of communication (pp. 10–11). Research on spoken interaction has suggested that non-verbal behavior is the least understood aspect, but at the same time acknowledges it as just as important as the words that are spoken (Knight, 2011; McNeill, 2016). This current study focuses particularly on the relationships between the co-occurrence of speech and gesture in English as a lingua franca (ELF) discourse by speakers of different first languages, across different English proficiency levels. It utilizes a corpus linguistics approach, with the aim of constructing a multimodal corpus of intercultural interaction and examining how such a corpus can be used to facilitate our understanding of gesture-speech relationship.

## 2. Speech and gestures

### 2.1. Sociocultural perspectives of speech and gestures

While language and gesture both play an important role in successful human interaction, they are regarded as being inseparable from culture; language, gesture, and culture are intricately interwoven such that each shapes the others. This is derived from the sociocultural theory (SCT) developed by psychologist L.S. Vygotsky (1986). SCT argues that human mental functioning is fundamentally a mediated process that is organized by cultural artifacts, including language and gesture, and culturally organized activities. In this regard, the process of becoming a human being is constituted through the process of being socialized into a particular culture (Negueruela and Lantolf, 2008). Gesturing styles among individuals of varied cultural groupings tend to differ based on form, space, and rate. Particularly, some cultures tend to be gesturally richer than others. Kendon (2004) illustrates this in his comparative study comparing Neapolitan gesturing to that by speakers in central England. He highlights the fact that Neapolitans are a gesturally rich culture, perhaps due to the public character of everyday life in Naples. Naples has historically been a densely populated city, indoor space was crowded, life was lived outdoors and, as a result, people made use of gesture for purposes of attention. The specific social and physical conditions for communication perhaps encouraged the use of gesture as an autonomous means of conversation, and this may explain why today Naples has such a rich gesture culture (Kendon, 2004, p. 353).

In addition, it has been reported that Italians use a great number of gestures in daily conversation (e.g., Kendon, 2004). Colletta et al. (2015) conducted a cross-linguistic investigation of multimodal narratives produced by French, American, and Italian children from two age groups, showing a higher proportion of representational gestures in the Italian children's repertoire than in the American children's, across both age groups. Previous research has also shown that the structure of the language itself influences gesture production, indicating that not all languages express space, location, and motion in the same way. For example, Özyürek et al. (2005) found that the gestural representations of the same events differ across languages when they are encoded by different syntactic frames (i.e., verb-framed or satellite-framed). Given that the role of gestures is crucial in human interaction, it is therefore important that the naturally occurring discourse in an intercultural

communication context is studied from a multimodal perspective. As Boxer (2002) notes, the study of discourse across cultures “represents an especially important endeavor in modern times” because of its great potential “for miscommunication and misperception based on differing norms of interaction across societies and speech communities” (p. 150). This study assesses this important area, in which differing gesture-speech patterning by participants from different language/culture backgrounds is examined.

## 2.2. *Communicative functions of L2 gestures*

Research on the role of gestures in foreign or second language (L2) speech has received increasing attention in recent years (Brown, 2015; Gregersen et al., 2009; Gullberg, 2006; Stam and McCafferty, 2008; So et al., 2013; Tutton, 2011). This suggests that the study of gesture-speech patterns in L2 speakers is considered be a valuable resource to better understand the connection of language, gesture, and communication. Previous research has shown the communicative functions of L2 gestures mainly taking on assistive roles, such as clarification, emphasis, avoidance of redundancy, replacement of speech, and seeking for help (Kendon, 2000, 2004; Zhao, 2006). Colletta et al. (2009, 2015) reported on 5- and 10-year-old children's use of gestures, which serve to reinforce, integrate, supplement, and complement their accompanying speech. As Kendon (2000) noted, “gestures, as used in partnership with speech, participate in the construction of the utterance's meaning. . . gestures can be used to provide context for spoken expression” (p. 60), which allows listeners to reduce dependence on their auditory skills and instead use visual cues in conversation, thus reducing the ambiguity of the meaning of what is expressed (Gregersen et al., 2009). Due to this partnership, gestures have the potential to serve as compensatory strategies either to facilitate lexical retrieval or elicit help from their interlocutors (Gullberg, 2006; Krauss, 1998). In L2 speech, speakers may employ multiple means to successfully get their information across, especially with limitations in speakers' proficiency levels. As Gullberg (2006) noted, gestures serve the function of bridging the communication gaps which result from lexical difficulties in communication.

L2 gesture researchers have also been interested in investigating whether the gesture-speech relationship and the way speakers gesture would differ between speakers at different L2 proficiency levels. Previous studies have reported a general tendency that the less proficient the speakers are in their L2, the more gestures they would produce when speaking in their L2 (e.g., Goldin-Meadow, 2005). This may be due to the fact that less proficient speakers employ gestures strategically to compensate for lexical problems or gaps in their L2 speech. Using gestures in speech in this regard allows them to compliment meaning in their verbal messages, which may otherwise be unintelligible without the accompanying gestures. Some studies, on the other hand, reported inconsistent findings, showing that gesture rates were significantly higher in advanced speakers than less proficient speakers (e.g., Gregersen et al., 2009; Gullberg, 1998). As Gregersen et al. (2009) noted, more proficient speakers use gestures in a meaning-enhancing way that allows them to emphasize specific information and to visually structure their verbal messages. This may, in turn, increase the probability of effective communication among interlocutors. Despite of the inconsistency in overall production of gestures, proficiency has been found to affect gesture production in complex ways.

## 2.3. *New generation corpus linguistics: multimodal corpora*

A multimodal corpus is defined as an annotated collection of “language and communication-related material drawing on more than one modality” (Allwood, 2008, p. 208). This adds an additional layer of complexity to corpus design and building in multimodal corpora involving not only text, as commonly found in traditional corpora, but other sensory (e.g., hearing) and production (e.g., gesture and speech) modalities (Allwood, 2008). The integration of textual, audial, and visual records of communicative events in multimodal corpora provides a platform for the exploration of a range of lexical, prosodic, and gestural features of conversation, and for investigation into the ways in which these features interact in authentic communication. Unlike monomodal corpora, which have a long history of use in linguistics, the construction and use of multimodal corpora is still in its relative infancy. Despite this, work using multimodal corpora has already proven invaluable for answering a variety of linguistic research questions that are otherwise difficult to consider.

One such multimodal corpus is the Nottingham Multi-Modal Corpus (NMMC), which presents its data in three different modes: textual, audial, and video records of naturally occurring interactions, aligning within a functional, searchable corpus interface. Access to lexical and gestural features of spoken discourse allows for the analysis of their interaction in the creation of authentic interaction. Analyzing the NMMC, Knight et al. (2009) explored the specific gesture sequences that are associated with discourse markers that function to manage the talk in academic supervision meetings. They also examined the extent to which head nods relate to language forms used to signal active listenership. These concerns of mapping verbal discourse function codes onto non-verbal gesture-in-talk is not possible to achieve when looking simply at textual transcriptions of communication. However, the fact that all data were collected from academic supervision meetings between Ph.D. students and their supervisors means that those specific behaviors observed are context or

genre dependent. Specific sequences of talk or gestures may be simply for the particular conversation, between those particular participants, or in that specific academic setting. Consequently, these behaviors may or may not be transferable beyond such a limited context.

Current multimodal corpora also tend to only feature a relatively small amount of data and/or focus on a specific discourse context, providing little utility for describing language use beyond this context. One example of this would be the Fruits Cart Corpus (Aist et al., 2006), which includes videos of task-oriented dialogs that are annotated for speaker's verbal requests and domain action executions. It consists of 104 digital videos of 13 participants, with each video ranging from 4 to 8 min in duration, resulting in a total of approximately 4000 utterances in the corpus. Other examples include the Mission Survival Corpus 1 (Mana et al., 2007), the NIST Meeting Room Phase II Corpus (Garofolo et al., 2004), and the VACE Multimodal Meeting Corpus (Chen et al., 2005). These are all multimodal corpora and each feature records of interactions extracted from one specific discourse context – a professional meeting room. In some of these corpora the content is scripted or pre-planned to a certain extent, and/or the conditions in which the recordings take place are controlled and experimental; for example, participants are told specifically what to discuss. Therefore, while the multimodal approach has proven effective in identifying patterns and relationships that traditional monomodal corpora cannot, these are still not necessarily appropriate for addressing research questions that focus on the more interpersonal aspects of communication, beyond a formal, professional contextual domain.

### 3. Research questions

While previous multimodal corpus research has extensively reported on gesture use in narratives (e.g., Gregersen et al., 2009; Holler and Wilkin, 2011; Parrill et al., 2010), English for specific contexts (e.g., Chen et al., 2005; Mana et al., 2007), academic contexts (e.g., Knight, 2011; Ovendale, 2012), and child language development (e.g., Colletta et al., 2009, 2015), few studies have addressed research questions focusing on more interpersonal and intercultural aspects of communication or consider the role of language proficiency of a speaker. This study aims to explore the patterns of co-occurrence of speech and gestures from an intercultural and multimodal corpus linguistics perspective, addressing the following research questions.

- *What semantic categories of speech most frequently co-occur with gestures? To what extent does the co-gesture speech differ among speakers from different cultural/linguistic backgrounds?*
- *What are the patterns of the gesture-speech relationship? To what extent does the relationship differ among speakers from different cultural/linguistic backgrounds and English proficiency levels?*

Co-gesture speech refers to the spoken words or phrases that are co-produced with hand gestures in face-to-face spoken conversation. These co-occurring words or entire lexical phrases were identified to reflect the meaning of the co-occurring gesture; they are also known as “lexical affiliates” of the gesture (Krauss, 1998). The first issue concerns the extent to which participants preferred to gesture particular semantic categories of information in their speaking. The second question looks at whether the gesture-speech relationship is to a certain extent influenced by language/culture backgrounds and English proficiency levels of a speaker. The examination of the co-occurrence of speech and gesture provides a key insight into the interplay between the two models of communication.

### 4. Methodology

#### 4.1. Corpus construction: MMCOIC

This research began with the development of the MMCOIC, Multimodal Corpus of Intercultural Communication. The participants recruited for this corpus development were 16 Taiwanese, 16 Indonesian, and 16 Indian students from a university in Taiwan. Gender ratio was equal for each group of the participants, and all participants were non-native English speakers, aged 20–25, having learned English as a second or foreign language for an average of eight to twelve years. Based on the CEFR,<sup>1</sup> the participants were equally distributed to advanced and low-intermediate levels according to their overall English proficiency. The advanced group include participants of upper-intermediate (B2+) to advanced (C1); the low-intermediate group mainly included students of low-intermediate (B1) level. Using the CEFR bands allows to the results of the study more replicable and comparable. The cross-group comparison also provides a larger cross section

<sup>1</sup> CEFR is an internationally recognised framework that describes different levels of language ability. For more information, see [http://www.coe.int/t/dg4/linguistic/cadre1\\_en.asp](http://www.coe.int/t/dg4/linguistic/cadre1_en.asp).

of participants of different L1 backgrounds and proficiency levels than a case study would and, potentially, a wider range of factors influencing the frequency, type, and functions of co-occurrence of speech and gesture.

During the data collection process, the participants were randomly paired up, with each pairing composed to two participants from different nationalities but with a similar English proficiency level. These paired participants met face-to-face and had a casual conversation in English in a university staff lounge. Each meeting lasted for 60 min. In the process of their conversation, the participants had the freedom to create and develop a conversational topic of common interest that sustained communication. They were not told the particular focus of the research before the conversation took place. All of the face-to-face meetings were video-recorded. The multimedia data collection resulted in approximately 24 h of recorded conversations, which were then transcribed into electronic written transcripts based on standard orthographic practices (VOICE Project, 2007), finally contributing to a 200,358-word corpus.

Throughout the process of data collection, great importance was placed on the naturally occurring nature of the discourse. By emphasizing this, it was believed that the credibility of the research results would be increased. However, it is difficult to promote real, naturalistic talk in research settings and, in this case especially, speakers may feel uneasy about being recorded due to the obtrusive nature of video cameras. This notwithstanding, as Adolphs and Carter (2013) noted, it is possible to obtain natural responses from participants, especially if the recordings take place in relaxed, familiar settings. The fact that each recording in the current study lasts approximately 1 h also means speakers have the chance to become more accustomed to their surroundings and the presence of the recording equipment. This, it is hoped, promoted talk that was as natural as possible. Additionally, the first 10 min of each conversation was not analyzed.

#### 4.2. Annotating co-gesture speech: semantic tagging

In order to explore the extent to which participants preferred to gesture particular categories of information in their speaking, the transcribed data needs to be annotated, namely each co-gesture lexical affiliate requires interpretative semantic information. To this end, all of the lexical affiliates in MMCOIC were collected in a plain text file, which was suitable for corpus analysis programs of semantic annotation. Uploading the text file to *Wmatrix* (Rayson, 2008; see also <http://ucl.ac.uk/lancs.ac.uk/Wmatrix>), a semantic tag was then automatically assigned to each of the lexical affiliates based on the USAS (UCREL Semantic Analysis System) tagger, containing 21 major discourse fields, which in turn expand into 232 fine-grained semantic categories (Rayson et al., 2004). Although *Wmatrix* provides users with a web-based tool to create and process the tagging automatically, Rayson (2008) reports that the accuracy rates of the semantic tagging are approximately 91%. As a result, while we can be confident in the tagging accuracy, the interpretation of the results should take into account possible errors. Manual checking of concordances with the video recordings was therefore done carefully in this current study.

#### 4.3. Coding co-speech gesture

While any movement of a body part that involves effort can be considered a gesture, this study focuses particularly on co-speech hand gestures, which can be defined as spontaneous communicative hand movements that accompany speech (McNeill, 2005). During the process of coding co-speech gestures, five coders were recruited and received extensive training. The transcripts were divided into two sets, and each set was analyzed together by two coders. The fifth coder then individually looked at 50% of the data from each set, validating annotations and settling any disagreements. The percentage of agreement for the this judgment was 98% for gesture type identification and 95% for classifying gesture-speech relationship.

For the gesture identification, each coder took into account the three following criteria based on Kendon (2004), with each criteria rated on a scale of 0 to 2, with 2 being the strongest value:

- (1) if the movement is easy to perceive, of good amplitude or marked well by its speed;
- (2) if location of the movement is in frontal space of the speaker;
- (3) if there is a precise hand shape or a well-marked trajectory.

Once the total score of each gesture stroke is equal to or above 3, the gesture was identified and annotated in the transcription. The coder then attributed a gesture type to each gesture stroke using McNeill's (1992, 2005) coding scheme, including deictic gestures, iconic gestures, and metaphoric gestures. Deictic gestures refer to objects or locations in physical or conceptual space. In face-to-face meetings, these gestures mostly are limited to pointing and often used in a reference to an imaginary placeholder. Deictic gestures are further divided into concrete and abstract deictic gestures because the referents can be concrete, such as the addressee, or they can be abstract, like pointing to the left and the right while uttering the words *good* and *bad*. Iconic gestures have close relationship to the semantic content of speech. They depict spatial or shape-oriented aspects of a referent, e.g. by using two fingers to indicate someone walking while uttering *he went down the street*. Iconic gestures provide visual cues and may further reinforce the verbal messages and enhance communication (McNeill, 2016). Metaphoric gestures are associated with abstract ideas. Similar to iconic



gestures in pictorial manifestation, they represent a metaphor of the speaker's idea or feeling about a concrete concept. Together, these categories of gesticulation provide a useful classification system based on how they function with the accompanying spoken message, as reported in previous research on co-speech gestures in a variety of communication contexts (Negueruela and Lantolf, 2008; Roth, 2001; So et al., 2013; Stam and McCafferty, 2008; Zhao, 2006). Although McNeill's (2016) categorization scheme also comprises other gesture types, such as adaptors, beats, and emblems, these were not included in this study. Adaptors and beats often do not encode semantic content of the speech. Emblems are gestures that are meaningful in and of themselves, and due to the fact that emblems typically replace speech instead of accompanying speech, they are not coded in this study.

#### 4.4. Identifying gesture-speech relationship

This investigation of the gesture-speech relationship employs a bottom-up approach, working first on the gestures simultaneously used with speech and then identifying their relation to the corresponding speech. The classification scheme, adapted from Colletta et al. (2009, 2015) and Kendon (2004), includes the following (with examples extracted from MMCOIC):

- (1) *Reinforcing*: the information brought by the gesture is identical to the linguistic information it is in relation with. The gesture resembles the physical properties and movement of objects or actions being described in speech, e.g., an iconic gesture in which the speaker was holding her own hair with the left hand and doing a scissor-like cutting motion to the hair with her right hand while saying *cut the hair* to reinforce the verbal message.
- (2) *Integrating*: the information provided by the gesture does not add supplementary information to the verbal message, but makes the abstract concepts more precise, e.g., a metaphoric gesture, pantomiming a spherical shape to represent the idea of wholeness, or an abstract deictic gesture, pointing to the left and the right with the whole left and right hand respectively while uttering the words *the good and the bad*.
- (3) *Supplementary*: the information brought by the gesture adds new information not coded in the linguistic content. For instance, while saying *she just left*, the speaker is using a deictic gesture by shifting the left index finger pointing to a specific direction repeatedly, indicating not only the displacement but also the direction of the displacement. Another example can be seen in an iconic gesture, in which the speaker used both hands to form a small round shape while saying *the teapot*. The interlocutor then replied *wow, so small*. In this case, the gesture depicts not only the object, but also the size, the information not included in verbal utterances.
- (4) *Complementary*: the information provided by the gesture brings a necessary complement to the incomplete linguistic information provided by the verbal message: the gesture disambiguates the message, as in the case of deixis (e.g., pointing gesture accompanying a location adverb like *here, there*; the pointing gesture identifying an object not explicitly named). It would not be clear what the deixis refers to without such a complementary gesture.
- (5) *Contradictory*: the information provided by the gesture contradicts the linguistic information provided by the verbal message. For example, while a Taiwanese speaker was saying *comes to Taiwan*, a gesture with two opening hands moving toward the hearer was produced.

The coders selected between *reinforcing*, *integrating*, *complementary*, *supplementary*, *contradictory*, and *others* if the gesture cannot be classified into the aforementioned categories. With regard to assigning a gesture to a particular category, it should be emphasized that gestures are typically densely-encoding and a gesture may serve multiple functions and can therefore be functionally ambiguous (Yoshioka, 2008). For example, as illustrated above a pointing gesture may at the same time perform more than one function, such as *integrating*, *complimenting* and even *contradictory*. As Tsui (1994) argues, the source of multiple functions often lies in the sequential environment of the conversation in which the utterance occurs (p. 45). As such, in this study the use of each gesture in its discourse context was examined to identify the primary function it serves in MMCOIC. When the gesture-speech relationship was still vague from the context, interviews with speakers were carried out to confirm whether the meanings of gestures had been interpreted appropriately. As suggested by Kochman et al. (2014), through multiple methods of data analysis, such as triangulation, we can check whether the interpretations of the results were both coherent and internally consistent.

## 5. Findings and discussion

### 5.1. Rates and frequencies of co-gesture speech

Table 1 presents the total words and clauses and the frequencies and rates of the co-gesture speech in MMCOIC, showing the percentages of lexical items that co-produced with gestures. It can be noted that the rates of co-gesture speech vary from 11.9% to 19.83%, showing that approximately 10–20% of the lexical items in speech were accompanied

Table 1  
Frequency of co-gesture speech in MMCOIC (distributed across L1 and proficiency groups).

	Taiwanese		Indian		Indonesian	
	Advanced	Low-intermediate	Advanced	Low-intermediate	Advanced	Low-intermediate
Total words	33,135	26,822	38,072	33,595	36,534	32,200
Clauses	5561	5314	5785	5788	6160	5805
Total words of co-gesture speech	3943	5318	4750	4401	4397	4395
Rates of co-gesture speech	11.90%	19.83%	12.48%	13.10%	12.04%	13.65%

with gestures. Previous empirical studies have shown that less proficient learners generally employed a greater number of non-verbal communication strategies in face-to-face foreign language communication than advanced speakers (e.g., Gregersen et al., 2009; Gullberg, 1998, 2006). However, in this current study an appreciable difference can only be seen in the Taiwanese dataset, and only slight differences were found in the Indian and Indonesian datasets with respect to the rates of co-gesture speech between advanced and low-intermediate-level speakers' utterances.

The face-to-face casual conversation among the participants includes a wide variety of topics, including traveling, food, hobbies, school life experiences, their own countries, etc. In order to explore the extent to which co-gesture speech differs among speakers from different cultural/linguistic backgrounds, Table 2 presents the most frequent lexical items co-occurring with gestures by different groups of participants in MMCOIC. While the wordlists are from three different groups, they are all highly similar, although the order of frequency in which they occur is slightly different. Clearly, the definite article *the* is the most frequent lexical item that accompanies with gesture in speaking by all groups of participants. Although in general *the* is used to refer to something which has already been mentioned (Carter and McCarthy, 2006), in MMCOIC *the* was frequently co-produced with gestures in speech where speakers define or identify a particular person or object, as in *the woman* and *the taste*. In addition, pronouns also frequently co-occurred with gestures. The large majority are first and second person pronouns (i.e., *I* and *you*, respectively), which encode the most basic aspect of a conversation between two people, the participants themselves (Lin, 2017; Petitto, 1987). Such high frequencies of first and second pronouns may be due to the fact that the conversations consisted of on only two interlocutors, and that gestures accompanying personal pronouns were formed by pointing directly to the addressee or self in the conversation.

Although the wordlists offers an immediate snapshot of what lexical items are co-produced with gestures in conversation, it is sometimes the case that some words do not occur often enough to be included in the high-frequency wordlist, and they therefore tend to be overlooked easily (Baker, 2006; Lin, 2015, 2017; Rayson, 2008). Baker (2006) exemplifies this with the notion of "largeness" in a text which is instantiated through various synonyms used by writers such as *big*, *large*, *huge*, *great*, *giant*, and *massive*. However, when none of these occur individually with high frequency, they would easily be neglected since the semantic similarities between words are not taken into account. In addition, wordlist usually generates far more lexical items than it is possible for the researcher to analyze (Lin, 2017; Rayson, 2008). Semantic domain analysis, therefore, would highlight the most significant categories that the researcher should examine. As a result, tagging the corpus with semantic domains could provide valuable information on the understanding of particular discourse features and thereby avoid the limitations of wordlists.

## 5.2. Semantic categories of co-gesture speech

*Wmatrix* was used to categorize all the co-gesture speech, or lexical affiliates to the related semantic domains. Frequencies of the semantic categories of the lexical affiliates and the overall speech were contrasted in order to reveal

Table 2  
Most frequent lexical items that co-occur with gestures.

	Taiwanese	Freq.	Indian	Freq.	Indonesian	Freq.
1	<b>THE</b>	3012	<b>THE</b>	3382	<b>THE</b>	3537
2	<b>YOU</b>	2432	<b>I</b>	2800	<b>I</b>	3001
3	<b>I</b>	2096	<b>YOU</b>	2438	<b>YOU</b>	2514
4	<b>IT</b>	1989	<b>A</b>	2348	<b>IT</b>	2414
5	<b>A</b>	1875	<b>IT</b>	2317	<b>A</b>	2248
6	<b>THEY</b>	1773	<b>IN</b>	2296	<b>IN</b>	2244
7	<b>IN</b>	1737	<b>THAT</b>	2261	<b>MY</b>	2058
8	<b>TO</b>	1731	<b>THEY</b>	2126	<b>TO</b>	2043
9	<b>WE</b>	1701	<b>MY</b>	2098	<b>THAT</b>	2039
10	<b>THIS</b>	1689	<b>HAVE</b>	2077	<b>WE</b>	2038

Table 3

Semantic domains that include highest proportion of co-gesture speech (with rates of speech co-occurring with gestures).

Rank	Semantic domain (Code)	Taiwanese Freq. (Rate)	Indian Freq. (Rate)	Indonesian Freq. (Rate)	Total (Rate)	Examples of common lexical affiliates
1	Moving, coming and going (M1)	1991 (71.1%)	2158 (69.4%)	2378 (73.4%)	6527 (71.3%)	<i>go, went, going, travel, come</i>
2	General objects (O2)	907 (74.2%)	1198 (67.8%)	1002 (71.2%)	3107 (71.0%)	<i>thing, things, machine, bowl, chopsticks</i>
3	Numbers (N1)	2320 (61.2%)	2575 (66.3%)	2023 (64.2%)	6918 (63.9%)	<i>one, two, ten, hundred, five</i>
4	Location and direction (M6)	2218 (58.8%)	4085 (52.0%)	4124 (53.7%)	10,427 (54.8%)	<i>this, here, there, where, that</i>
5	Time (T1)	2042 (49.2%)	2201 (53.8%)	2196 (54.1%)	6439 (52.4%)	<i>time, days, year, night, week</i>

the categories that exhibit a high proportion of co-gesture speech. Table 3 illustrates the five semantic domains that include highest proportion of co-gesture speech by Taiwanese, Indian, and Indonesian participants respectively with rates of speech that co-occur with gestures. For example, overall 71.3% of the speech expressing moving, coming and going (M1) was co-produced with gestures. From this analysis, we can see how co-gesture speech was produced differently by participants of different linguistic backgrounds. It can be seen that all groups of participants' co-gesture speech most commonly occurs with location and direction (e.g., *this, here*), moving, coming and going (e.g., *go, come*), general objects (e.g., *thing, machine*), numbers (e.g., *one, two*), and periods of time (e.g., *days, week*). Other semantic domains exhibiting high proportion of co-gesture speech include geographical names (e.g., *Taiwan, Taipei*), people (e.g., *person, kid*), general actions (e.g., *do, make*), obtaining and possession (e.g., *get, take*), and pronouns (e.g., *you, me*). Although participants of different linguistic backgrounds tended to gesture in their speech in different semantic domains, the top five domains are exactly the same among the three different groups (Taiwanese, Indian, Indonesian). This indicates similar preferences of gesture-speech production by speakers despite the different cultural and linguistic backgrounds.

Among these semantic domains of co-gesture speech, it can be seen that many of the lexical items are deictic expressions, referring to specific entities, people, places, etc., as in the categories of general objects (O2), location and direction (M6), people (S2), pronouns (Z8), and geographical names (Z2). The lexical items within these semantic domains were commonly accompanied with deictic gestures (i.e. gestures that point to physical entities, directions, or a location), which identify referents in the gesture space in order to insure that the intended message is effectively conveyed by speech and gesture (Kita and Özyürek, 2003). For example, the demonstratives (i.e. *this, that*) are pervasive in co-gesture speech in MMCOIC with a typical function of identifying concrete referents in space. In a similar vein, geographical names (Z2), such as *Taiwan, India, and Indonesia*, involving the countries where the participants come from, were also commonly produced as co-gesture speech. As Lin (2015) noted, in intercultural interaction participants talk about their own cultures and contrast the differences among different countries; at the same time, they demonstrate a willingness to learn from other international peers, showing interest in language, culture, and life experiences. Gesture, in this regard, was simultaneously used with speech in order to indicate the countries and the differences among them. As has been argued earlier, gestures and language are intrinsically linked. The findings here further support previous research on the collaborative use of L2 speech and gesture, showing that L2 speakers (compared to L1) tended to frequently position people, objects, or events in their discourse in space through gesture and refer back to them using deictic gestures (Gullberg, 2006; Stam and McCafferty, 2008).

Moreover, the speech in the semantic domains of moving, coming and going (M1), general actions (A1.1), and obtaining and possession (A9+) also frequently co-occurred with gestures. Examination of the co-occurring gestures indicates that relatively more iconic gestures were found in these semantic domains. Iconic gestures depict actions or motions, such as right hand moving forward and backward with palm facing the listeners and the speaker respectively in turn to convey *going* and *coming*, or the right hand acting out as if catching/getting something or moving something from where the speaker is to a different place to indicate *get* or *take*. These spontaneous iconic gestures depicting action or perceptual information provide visual cues and thereby may reinforce the verbal messages and enhance communication. In this regard, as noted by McNeill (2005), speech and gesture were “co-expressive” in the sense that “they portrayed the same event but were not redundant, since each articulated its own aspect of it” (p. 39). Previous studies have also indicated the high-frequency co-occurrence of iconic gestures with verbs or verb phrases (e.g., Capone and McGregor, 2004). In MMCOIC, the lexical items within these two semantic domains include a large portion of verbs (e.g., *go, take, get*).



### 5.3. Gesture-speech relationship

The relationship between speech and its accompanying gestures investigated here fall into six categories: *reinforcing*, *integrating*, *supplementary*, *complementary*, *contradictory*, and others. With a view to making the findings comparable across groups, the total number of speech-accompanying gestures are divided by the number of clauses, as suggested by Colletta et al. (2015). Table 4 displays the mean frequencies of gestures per 10 clauses (with relative proportions across types of relationship) and the distribution of the different gesture-speech relationships by Taiwanese, Indonesian, and Indian participants of two different proficiency levels. All the data were analyzed using one-way ANOVA with independent samples. Language/Culture backgrounds (Taiwanese, Indonesian, and Indian) and English proficiency levels (advanced and low-intermediate) were regarded as between-groups factors.

In the analysis of mean frequencies of total types of gesture-speech relationships, only proficiency level was found to have a significant effect ( $F(1, 46) = 15.32, p = 0.00$ ). That is, the gesture-speech relationship was not significantly influenced by different language backgrounds of speakers ( $F(2, 45) = 2.53, p = 0.21$ ). This notwithstanding, the post hoc comparison using the Scheffe test revealed that Taiwanese speakers tended to use significantly more complementary gestures but less reinforcing gestures ( $p = 0.013$  and  $0.002$ , respectively) compared with Indonesian speakers. On the other hand, comparing the gesture-speech relationships by participants between advanced and low-intermediate levels, the more proficient speakers from all three different language groups tended to use significantly more gestures that serve to reinforce and integrate ( $F(1, 46) = 96.58, p = 0.00$  and  $F(1, 46) = 22.92, p = 0.001$ , respectively), while low-intermediate-level speakers employed more gestures as *complements* and *others* ( $F(1, 46) = 69.03, p = 0.00$  and  $F(1, 46) = 94.50, p = 0.00$ , respectively). The following will discuss each gesture-speech relationship in detail.

As can be seen in Table 4, *reinforcing* is the most frequent type of gesture-speech relationship; all three language groups of the participants across different proficiency levels most frequently reinforce the linguistic information in their speech through gestures. As McNeill et al. (2000) suggest, in many situations when a gesture accompanies a verbal message it often reinforces the accompanying speech, in the sense that they convey the same semantic content as speech. Speakers in this regard move their hands and arms to create a dynamic visual representation of semantic properties related to the content of their speech. Such gestures display an aspect of the same scene that speech also presents, as in excerpt (1):

- (1) <ID01>: Is there any special food that you are very best at it in cooking in the food? Or for maybe you want to offer me. . . OK. . . I want to cook you  
 <TW01>: **Hotpot** [*Reinforcing*]  
 <ID01>: Because you can only **put** [*Reinforcing*]  
 <TW01>: Yeah. You also don't need to **fry** [*Reinforcing*], you don't need to **cut** [*Reinforcing*], you just **put** [*Reinforcing*] everything in the hotpot

Five gesture strokes that served as *reinforces* were identified in (1). These strokes of gestures are synchronized with the co-expressive words *hotpot* (both hands to form a circle as a big pot), *put* (right hand moving vertically as if putting something in the pot), *fry* (right hand swinging, acting out as if frying something), and *cut* (tightly closed fingers of right hand, acting out as a kitchen knife chopping something), respectively. In this regard, speakers create a dynamic visual representation identical to the content of the verbal message, and this helps listeners to build an enriched conceptual representation of the speaker's message (McNeill, 2005, 2016; Kendon, 2004). Such gestures can be "a means of mapping intentionality onto lexical items" in order to facilitate listener's comprehension (Stam and McCafferty, 2008, p. 12). What needs to be noted is that the ANOVA analysis shows that advanced speakers particularly favored this gesture-speech relationship, which may be due to the fact that proficient foreign language speakers tend to use hand gestures accompanied by verbal messages to

Table 4  
 Gesture-speech relationship in MMCOIC: Mean frequencies (with relative proportions across types of relationship).

Participants	Total	Reinforcing	Integrating	Supplementary	Complementary	Contradictory	Other
Taiwanese							
Advanced	7.27	2.53 (35%)	1.63 (22%)	2.22 (31%)	0.67 (9%)	0.03 (0%)	0.19 (3%)
Low-intermediate	6.50	1.66 (25%)	0.92 (14%)	1.91 (29%)	1.54 (24%)	0.02 (0%)	0.45 (7%)
Indian							
Advanced	7.97	3.02 (38%)	1.87 (23%)	2.12 (27%)	0.69 (9%)	0.02 (0%)	0.25 (3%)
Low-intermediate	6.59	2.01 (31%)	1.23 (19%)	1.86 (28%)	1.18 (18%)	0.03 (0%)	0.28 (4%)
Indonesian							
Advanced	7.50	3.14 (42%)	1.59 (21%)	2.03 (27%)	0.52 (7%)	0.03 (0%)	0.19 (3%)
Low-intermediate	6.19	1.88 (30%)	1.05 (17%)	2.19 (35%)	0.77 (12%)	0.03 (0%)	0.27 (4%)

co-construct meaning, which may well further facilitate communication (Gregersen et al., 2009). Whereas reinforcing gestures resemble concrete objects or actions, in this study the gestures which not only provide equivalent visual representation but also add precision to something abstract was coded as *integrating*.

*Integrating* is the second major function served by co-speech gestures in MMCOIC. From the ANOVA analysis, this gesture-speech relationship was shown to be another critical feature in distinguishing gestures from low and more proficient speakers. The information provided by such gestures adds precision to the abstract concepts of the linguistic information. An example of integrating gestures includes doing a finger quote with the index and middle fingers on both hands as a standard for a signal for a phrase or a quote of an abstract concept, or a metaphoric gesture, which is typically conceived of as movements of the hands that represent or indicate the source domain of a metaphor (McNeill, 2005). In (2) for example speaker IN03 made a fist with her right hand and beat four times on her open left hand, visualizing the abstract concept of the amount of pressure.

- (2) <IN03>: You know my friends my friends told me uh you want to take your PhD program, you need to think twice. Because they got **so much pressure** [Integrating] from their professor, so...
- (3) <ID02>: For example, **2005** [Integrating], **2015** [Integrating] **10 years' time** [Integrating] **I started** [Integrating], reading everything like a long study...

In (3), while ID02 was talking about her experiences of working as a research assistant in the past 10 years, she used her left hand (pointing to the left) to indicate 2005 and her right hand (pointing to the right) to indicate 2015, and then two hands moving closer to each other while saying *10 years' time* to represent the duration of the 10 years between 2005 and 2015. In the following utterance, the gesture adds precision to the linguistic information: the speaker said *I started*, and at the same time she points to her left, which indicates the time she started being a research assistant. The speaker's gestures (but not her words) make it clear that she started from 2002 even the time is not mentioned again in speaking. In this case, the speech-accompanying gestures were in abstract discourse-internal uses (with abstract deictic gesture) where the meaning is dependent on something other than the relative physical location of the speaker (Fromkin, 2016; Johansson and Zlatev, 2013).

Third, the information brought by the *supplementary* gesture adds new information not coded in the linguistic content due to the fact that information brought by the representational gesture may say more than the linguistic information (McNeill, 2005; Kendon, 2004). One such is in (4) when the speaker IN04 tries to describe a lunch box by demonstrating a tridimensional octagon in the air, the gesture provides additional information to show the shape and the size of the lunch box, which is not included in verbal messages. Another example can be seen in (5). While TW07 was uttering *Okay*, his left hand pointed to IN07, attempting the interlocutor to continue saying the things he remembered. In this case, the gesture conveyed an additional meaning, similar to *Can you talk about it?*, which adds new information to the speech *Okay*. With regard to the use of *supplementary* gesture, no significant difference was found among participants from different language backgrounds or proficiency levels.

- (4) <TW04>: You can try the lunch box on the train.  
<IN04>: I know it's famous. Is it like a **lunch box** [Supplementary] and...  
<TW04>: Yes yes it's like that, octagon.
- (5) <IN07>: Oh! I, I, I...I remember something.  
<TW07>: **Okay** [Supplementary].  
<IN07>: Yeah, it's ab- about...

Fourth, in the analysis of *complementary* gestures, proficiency level was found to have a significant effect, showing that the speakers at low-intermediate level tended to produce significantly more co-gesture speech as *complements*. The information provided by such gesture brings a necessary complement to the incomplete linguistic information provided by the verbal message, so the gesture helps to disambiguate.

- (6) <TW06> So, in India, **the situation** [Complementary] is the same?  
<IN06> Yeah I think it's the same. Because right now I think it's also around 6 pm. It's getting dark. Not this dark I think. Because... I think...I mean it's very dark now...I mean India around 6 p.m. at this time now.
- (7) <ID07> I've never been there.  
<TW07> **Here** [Complementary]. **This** [Complementary] is so beautiful.

In (6), TW06 and IN06 were discussing the weather outside the room. When a pointing gesture toward the sky accompanies the incomplete linguistic information in speech (i.e. *the situation*), the gesture disambiguates the message. It would not be clear what *the situation* meant if no gesture was employed. A similar example can be seen in (7), in which TW07 showed a picture from his smartphone, and pointing gestures accompanied brings necessary complements to *here* and *this*. In this case, the speech-accompanying gestures were used in concrete spatial deixis, incorporating the context of the physical surroundings of the interlocutors (Fromkin, 2016). Due to the fact that such gestures are driven by physically present referents that are perceptually accessible to both the speaker and listener, they provide an ideal vehicle to meet communicative goals in face-to-face conversation (So et al., 2013). In this regard, the speakers do not need to lexically specify the referent, co-occurring concrete deictic gestures were instead used to perform the task of specifying. Such a phenomenon of less proficient speakers produced significantly more concrete deictic gestures than advanced speakers is supported by previous studies on co-speech gestures by L2 speakers (e.g., Negueruela and Lantolf, 2008; So et al., 2009, 2013).

The fifth type of gesture-speech relationship is *contradictory*. In some cases, while mentioning *you*, a speaker pointed to him/herself, which may indicate *me*. Another example is shown in (8), in which the Taiwanese speaker was doing a beckoning/inviting hand gesture while saying *back to India*. The information provided by the gesture (i.e. *coming*) contradicts the linguistic information provided by the verbal message (i.e. *back to*). This happens to both advanced and low-intermediate speakers, whereas the frequencies are the least among different gesture-speech relationship.

- (8) < TW01> So don't you think about **back to** [Contradictory] India to see your parents.  
 < ID01> Actually that will be a very good thing to do. But the thing is. . .

Last but not least, as can be seen in Table 4 a number of gestures were identified as “Other”. This category includes the gestures which have no obvious relationship to the conceptual content of their accompanying speech. The comparison of the rates of such gesture-speech relationships shows a significant difference between advanced and low-intermediate-level participants. This is possibly due to the fact that speakers with lower levels of language proficiency may not be able to process information quickly enough in speaking and further use gestures to enhance the meaning of their verbal messages. Another possible reason may be due to foreign language anxiety. For example, failures of word retrieval may be frustrating, and this frustration-generated tension could interfere with the speaker's gesture use and ability to produce coherent speech. As has been noted by previous studies (Gregersen, 2005; Gregersen et al., 2009), anxious learners tend to use their hands for less illustrative and regulatory purposes, and more for fidgeting and interactionally nonproductive activity under the stress of using their L2 in a speaking situation. In MMCOIC, speakers at higher levels of language proficiency used more speech-related gestures that served a clearer function to the accompanying speech.

## 6. Conclusion

This present study investigates the intercultural interaction among Taiwanese, Indonesian, and Indian speakers utilizing a multimodal corpus linguistics approach based on MMCOIC with a view to examining how co-gesture speech is produced at semantic levels and whether the gesture-speech relationship is to a certain extent influenced by language/culture backgrounds and English proficiency levels of a speaker. The semantic domains of the co-gesture speech demonstrate not only different preferences to gesture particular semantic categories of information in their speaking by the participants, but also similarities and the amount of overlap among participants of different language/culture backgrounds and English proficiency levels. In all three groups of speakers, for example, co-gesture speech most commonly occurs in general objects (e.g., *thing*, *machine*), moving, coming and going (e.g., *go*, *come*), location and direction (e.g., *this*, *here*), numbers (e.g., *one*, *two*), and periods of time (e.g., *days*, *week*). These high-frequency semantic categories of the co-gesture speech were mainly referring expressions accompanying with deictic gestures, which identify referents in the gesture space. Such pervasive use of deictic gestures can be seen in a variety of communication contexts, as reported in previous research on co-gesture speech (Allen, 2003; Colletta et al., 2009, 2015; So et al., 2013; Stam and McCafferty, 2008).

The gesture-speech relationship falls into six categories: *reinforcing*, *integrating*, *supplementary*, *complementary*, *contradictory*, and *others*. While the gesture-speech relationship was not significantly influenced by different language backgrounds of the speaker, more proficient speakers from all three different language groups tended to use significantly more gestures that serve *reinforcing* and *integrating* functions; on the other hand, low-intermediate-level speakers produced more gestures as *complements* and other gestures that have no obvious relationship to the conceptual content of their accompanying speech. The high frequencies of reinforcing and integrating gestures probably occur as the result of a listener-oriented rather than message-oriented strategy (Adolphs and Carter, 2013), where more proficient speakers highlight the semantic contents of speech to ensure clarity and may further facilitate communication. To summarize, this study provides evidence to support the idea that there exists similar preferences of gesture-speech production by

speakers despite their different language/culture backgrounds, whereas the gesture-speech relationship differs between speakers of high and low proficiency levels.

Although this paper has explored the co-occurrence of speech and gesture in an intercultural setting, due to the small numbers of participants of different L1s and additionally the fact that the conversation setting during data collection was between only two participants from different Asian countries in casual conversation, the results from this sample should be interpreted with caution before generalizing to intercultural interaction as a whole. This notwithstanding, the size and composition of the MMCOIC makes it more manageable for multimodal studies and permits a closer link between language and gestures in context to understand the semantic and pragmatic aspects regarding the collaborative use of speech and gesture.

This study also has important pedagogical implications. It has long been understood that language itself is inseparable from gesture, and therefore, it may not be possible for a foreign language learner to become a successful user of a language without also learning the appropriate use of gestures. The study further suggests that learners at different levels of English proficiency could enhance the meaning of verbal messages trying to be conveyed by the speaker through speech-related gestures to reinforce, integrate, compliment, and supplement their speech. Learners at lower proficiency levels could benefit from training that includes using gestures to reinforce and integrate their spoken language, while more proficient speakers, as has been suggested (e.g., Gregersen et al., 2009; So et al., 2009, 2013), could learn to reduce the ambiguity of their verbal messages by presenting their speech using multimodal channels of communication. Further investigations on the teaching of “gestural competence” (Gregersen et al., 2009, p. 206) would be of interest for a better understanding of gesture-speech system development in foreign or second language acquisition.

## References

- Adolphs, S., Carter, R., 2013. *Spoken Corpus Linguistics: From Monomodal to Multimodal*. Routledge, London.
- Aist, G., Allen, J., Campana, E., Galescu, L., Gómez Gallo, C., Stoness, S., Swift, M., Tanenhaus, M., 2006. Software architectures for incremental understanding of human speech. In: *Proceedings of Interspeech 2006*, pp. 1922–1925.
- Alibali, M., Hostetter, A., Kita, S., 2007. I see it in my hands' eye: representational gestures reflect conceptual demands. *Lang. Cogn. Process.* 22 (3), 313–336.
- Allen, G.L., 2003. Gestures accompanying verbal route directions: do they point to a new avenue for examining spatial representations? *Spatial Cogn. Comput.* 3 (4), 259–268.
- Allwood, J., 2008. Multimodal corpora. In: Lüdeling, A., Kytö, M. (Eds.), *Corpus Linguistics: An International Handbook*, vol. 1. Mouton de Gruyter, Berlin, pp. 207–225.
- Arslangul, A., 2015. How French learners of Chinese L2 express motion events in narratives. In: Su, D., Fu, J. (Eds.), *Space and Quantification in Languages of China*. Springer International Publishing, pp. 165–187.
- Baker, P., 2006. *Using Corpora in Discourse Analysis*. Continuum, London.
- Boxer, D., 2002. Discourse issues in cross-cultural pragmatics. *Ann. Rev. Appl. Linguist.* 22, 150–167.
- Brown, A., 2015. Universal development and L1–L2 convergence in bilingual construal of manner in speech and gesture in Mandarin, Japanese, and English. *Mod. Lang. J.* 99 (S1), 66–82.
- Butcher, C., Goldin-Meadow, S., 2000. Gesture and the transition from one- to two-word speech: when hand and mouth come together. In: McNeill, D. (Ed.), *Language and Gesture*. Cambridge University Press, Cambridge, pp. 141–161.
- Canale, M., 1983. From communicative competence to communicative language pedagogy. In: Richards, J.C., Schmidt, R.W. (Eds.), *Language and Communication*. Longman, New York, pp. 2–27.
- Capone, N.C., McGregor, K.K., 2004. Gesture development: a review for clinical and research practices. *J. Speech Lang. Hear. Res.* 47 (1), 173–186.
- Carter, R., McCarthy, M., 2006. *Cambridge Grammar of English: A Comprehensive Guide*. Cambridge University Press, Cambridge.
- Chen, L., Travis-Rose, R., Parrill, F., Han, X., Tu, J., Huang, Z., Harper, M., Quek, F., McNeill, D., Tuttle, R., Huang, T., 2005. VACE multimodal meeting corpus. In: *Proceedings of the Workshop on Machine Learning for Multimodal Interaction (MLMI)*. Retrieved from: [http://groups.inf.ed.ac.uk/mlmi05/tech\\_prog/arch/M-B-1.pdf](http://groups.inf.ed.ac.uk/mlmi05/tech_prog/arch/M-B-1.pdf)
- Colletta, J.M., Kunene, R.N., Venouil, A., Kaufmann, V., Simon, J.P., 2009. Multi-track annotation of child language and gestures. In: Kipp, M., Martin, J.C., Paggio, P., Heylen, D. (Eds.), *Multimodal Corpora: From Models of Natural Interactions to Systems and Applications*. Springer, Berlin, pp. 54–72.
- Colletta, J.M., Guidetti, M., Capirci, O., Cristilli, C., Demir, O.E., Kunene-Nicolas, R., Levine, S., 2015. Effects of age and language on co-speech gesture production: an investigation of French, American, and Italian children's narratives. *J. Child Lang.* 42 (1), 122–145.
- Fromkin, V., 2016. *An Introduction to Language*, 10th ed. Content Technologies, Inc., CA.
- Garofolo, J., Laprun, C., Michel, M., Stanford, V., Tabassi, E., 2004. The NIST meeting room pilot corpus. In: *Proceedings of the 4th Language Resources and Evaluation Conference (LREC)*.
- Goldin-Meadow, S., 2005. *Hearing Gesture: How Our Hands Help Us Think*. Harvard University Press.
- Gregersen, T.S., 2005. Nonverbal cues: clues to the detection of foreign language anxiety. *Foreign Lang. Ann.* 38 (3), 388–400.
- Gregersen, T., Olivares-Cuhat, G., Storm, J., 2009. An examination of L1 and L2 gesture use: what role does proficiency play? *Mod. Lang. J.* 93 (2), 195–208.
- Gullberg, M., 1998. *Gesture as a Communication Strategy in Second Language Discourse: A Study of Learners of French and Swedish*. Lund University Press, Lund, Sweden.

- Gullberg, M., 2006. Some reasons for studying gesture and second language acquisition (Homage à Adam Kendon). *IRAL Int. Rev. Appl. Linguist. Lang. Teach.* 44 (2), 103–124.
- Holler, J., Wilkin, K., 2011. An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers' responses. *J. Pragmat.* 43 (14), 3522–3536.
- Johansson, N., Zlatev, J., 2013. Motivations for sound symbolism in spatial deixis: a typological study of 101 languages. *Public J. Semiot.* 5 (1), 3–20.
- Kendon, A., 2000. Language and gesture: unity or duality. In: McNeill, D. (Ed.), *Language and Gesture: Window into Thought and Action*. Cambridge University Press, Cambridge, pp. 47–63.
- Kendon, A., 2004. *Gesture: Visible Action as Utterance*. Cambridge University Press, Cambridge.
- 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. *J. Mem. Lang.* 48 (1), 16–32.
- Knight, D., 2011. *Multimodality and Active Listenership: A Corpus Approach: Corpus and Discourse*. Bloomsbury, London.
- Knight, D., Evans, D., Carter, R., Adolphs, S., 2009. HeadTalk, HandTalk and the corpus: towards a framework for multi-modal, multi-media corpus development. *Corpora* 4 (1), 1–32.
- Kochman, K., Moelants, D., Leman, M., 2014. Gesture as a communicative tool in vocal pedagogy. *J. Interdiscip. Music Stud.* 6 (2), 233–250.
- Krauss, R.M., 1998. Why do we gesture when we speak? *Curr. Dir. Psychol. Sci.* 7 (2), 54–60.
- Lin, Y.L., 2015. Using key part-of-speech analysis to examine spoken discourse by Taiwanese EFL learners. *ReCALL* 27 (3), 304–320.
- Lin, Y.L., 2017. Keywords, semantic domains and intercultural competence in the British and Taiwanese Teenage Intercultural Communication Corpus. *Corpora* 12 (3).
- Mana, N., Lepri, B., Chippendale, P., Cappelletti, A., Pianesi, F., Svaizer, P., Zancanaro, M., 2007. Multimodal corpus of multi-party meetings for automatic social behavior analysis and personality traits detection. In: Mase, K., Massaro, D. (Eds.), *Proceeding of Workshop on Tagging, Mining and Retrieval of Human-related Activity Information at ICM'07*. ACM, New York, pp. 9–14.
- Mayberry, R.I., Jaques, J., 2000. Gesture production during stuttered speech: insights into the nature of gesture-speech integration. In: McNeill, D. (Ed.), *Language and Gesture*. Cambridge University Press, Cambridge, pp. 199–214.
- McNeil, N.M., Alibali, M.W., Evans, J.L., 2000. The role of gesture in children's comprehension of spoken language: now they need it, now they don't. *J. Nonverb. Behav.* 24 (2), 131–150.
- McNeill, D., 1992. *Hand and Mind: What Gestures Reveal About Thought*. University of Chicago Press, Chicago.
- McNeill, D., 2005. *Gesture & Thought*. The University of Chicago Press, Chicago.
- McNeill, D., 2016. *Why We Gesture: The Surprising Role of Hand Movement in Communication*. Cambridge University Press, Cambridge.
- Morsella, E., Krauss, R.M., 2004. The role of gestures in spatial working memory and speech. *Am. J. Psychol.* 117, 411–424.
- Negueruela, E., Lantolf, J.P., 2008. The dialectics of gesture in the construction of meaning in second language oral narratives. In: McCafferty, S., Stam, G. (Eds.), *Gesture: Second Language Acquisition and Classroom Research*. Routledge, NY, pp. 88–106.
- Ovendale, A., 2012. *The Role of Gesture in Cross-cultural and Cross-linguistic Learning Contexts: The Effect of Gesture on the Learning of Mathematics* (Unpublished doctoral dissertation). University of Johannesburg, South Africa.
- Özyürek, A., Kita, S., Allen, S., Furman, R., Brown, A., 2005. How does linguistic framing of events influence co-speech gestures? Insights from cross-linguistic variations and similarities. *Gesture* 5, 219–240.
- Parrill, F., Bullen, J., Hoburg, H., 2010. Effects of input modality on speech-gesture integration. *J. Pragmat.* 42 (11), 3130–3137.
- Petitto, L.A., 1987. On the autonomy of language and gesture: evidence from the acquisition of personal pronouns in American Sign Language. *Cognition* 27 (1), 1–52.
- Rayson, P., 2008. From key words to key semantic domains. *Int. J. Corpus Linguist.* 13 (4), 519–549.
- Rayson, P., Archer, D., Piao, S.L., McEnery, T., 2004. The UCREL semantic analysis system. In: *Proceedings of 4th International Conference on Language Resources and Evaluation (LREC, 2004)*, pp. 7–12.
- Roth, W.M., 2001. Gestures: their role in teaching and learning. *Rev. Educ. Res.* 71 (3), 365–392.
- So, W.C., Kita, S., Goldin-Meadow, S., 2009. Using the hands to identify who does what to whom: gesture and speech go hand-in-hand. *Cogn. Sci.* 33 (1), 115–125.
- 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. *J. Psycholinguist. Res.* 42 (6), 581–594.
- Stam, G., McCafferty, S., 2008. Gesture studies and second language acquisition: a review. In: McCafferty, S., Stam, G. (Eds.), *Gesture: Second Language Acquisition and Classroom Research*. Routledge, NY, pp. 3–24.
- Tsui, A.B.M., 1994. *English Conversation*. Oxford University Press, Oxford.
- Tutton, M., 2011. How speakers gesture when encoding location with English on and French sur. *J. Pragmat.* 43 (14), 3431–3454.
- VOICE Project, 2007. *VOICE Transcription Conventions* [2.1]. Retrieved from: [http://www.univie.ac.at/voice/voice.php?page=transcription\\_general\\_information](http://www.univie.ac.at/voice/voice.php?page=transcription_general_information)
- Vygotsky, L.S., 1986. *Thought and Language*. MIT Press, Cambridge, MA.
- Yoshioka, K., 2008. Linguistic and gestural introduction of Ground reference in L1 and L2 narrative. In: McCafferty, S., Stam, G. (Eds.), *Gesture: Second Language Acquisition and Classroom Research*. Routledge, NY, pp. 211–230.
- Zhao, J., 2006. The communicative functions of gestures in L2 speech. *Ariz. Work. Pap. SLAT* 13, 1–17.

**Yen-Liang Lin** is currently an Assistant Professor in the Department of English at National Taipei University of Technology, Taiwan. He received his Ph.D. in Applied Linguistics in 2013 from the University of Nottingham, UK. He has taught EFL at secondary and university levels for more than 10 years. His research interests include corpus linguistics, speech and gesture, intercultural interaction, computer-mediated communication and language teaching methodology.