Cross-Modal Grounding of Meaning

Kawai Chui (kawai@nccu.edu.tw)
Department of English & Research Center for Mind, Brain and Learning, No. 64, Sec. 2, ZhiNan Rd.
Mucha, Taipei City, 11605, Taiwan

Abstract

The current study investigates how the speaker grounds meaning for a referent by gestural repetition along with speech in daily conversation. The domain of analysis is the stretch of talk that encompasses the beginning and the end of the joint action during which a pair of similar gestures is produced by different speakers across turns to depict the same referent. A particular cross-modal grounding strategy was found: The previous speaker's gesture is replicated to form a semantic foundation shared by the participants; upon such common ground in gesture, the speaker readily conveys a new meaning with a new lexical expression. The use of the cross-modal grounding strategy facilitates the expression of shared knowledge in gesture and new meaning in speech within a clausal unit. It also bears out the bilateral process of speaking: During the construction of utterance, the speaker grounds new meaning by considering the addressee's current state of knowledge as revealed by gestural repetition. In turn, the addressee informs the speaker about his/her acknowledgement of the newly established meaning.

Keywords: grounding; meaning; gestural repetition; cross-modal strategy

Introduction

Grounding is an important aspect of language use (Clark, 1996; Clark & Brennan, 1991; Clark & Krych, 2004; Clark & Schaefer, 1987, 1989; Clark & Wilkes-Gibbs, 1986). "In dialogue, speakers try to ground their communicative acts as they go along: They work with their partners to reach the mutual belief that the partners have understood them well enough for current purposes" (Clark & Krych, 2004: 63). In Clark and Krych's (2004) study, pairs of participants, each comprised of a Director and a Builder, engaged in the task of assembling ten Lego models. It was found that whether the Director could see the Builder or not, and whether both were given instructions by audiotape or not affected the linguistic and gestural performances of the participants in grounding the Lego pieces. The findings provided evidence that being able to see the addressees' workspaces was crucial to making grounding more efficient. In daily conversation, participants also see each other. Then, when they do not engage in referent communication tasks, how do participants ground meaning by speech and gesture? The present study investigates participants' linguistic and gestural manifestations in the act of the grounding of meaning for a referent under discussion in conversation.

While multimodal resources can be employed to accomplish the social communicative act of providing semantic information for mutual understanding between the speaker and the addressee(s), the study focuses on gestural repetition along with speech. The re-enactment of a manual configuration which has previously been produced by another speaker for the same referent is called 'gestural mimicry' (Kimbara, 2006), and gestures of this kind are called 'return gestures' (de Fornel, 1992), 'gestural rephrasings' (Tabensky, 2001), 'mimicking gestures' (Kimbara, 2006) or 'mimicked gestures' (Holler & Wilkin, 2011; Parrill & Kimbara, 2006). The current study aims to investigate gestural repetition along with speech in conversational interaction during which speakers provide and ground meaning naturally. It will be demonstrated that gestural repetition can reveal the interactive nature of grounding, and provide the clearest evidence for speaking as a bilateral process (Clark & Krych, 2004), manifesting what information about addressees' state of knowledge speakers would take into account and how speakers use that information during the construction of their linguistic-gestural units, and also how addressees inform speakers about their understanding.

Data and Method

The study was based on videotaped data from daily face-to-face conversations in Mandarin Chinese. The participants were recruited to hold a conversation with their friends, family members or colleagues who knew each other. For every recording, the participants chose a place where they could talk in a leisurely manner, such as a classroom, students' lounge, dorm room or living room. Participants were free to find and develop topics of common interest, and were filmed for approximately an hour with a visible camera. All of the participants were paid, and they were told that they were participating in a study on conversation; gestures were not mentioned. One stretch from each talk, of about twenty to forty minutes, in which the participants were comfortable in front of the camera, was then selected for transcription. The speech and the gesture data relevant for the present study were extracts from seven conversations among adult native speakers of Mandarin for a total of about 140 minutes of talk; seven males and ten females were involved.

The domain of analysis is 'a stretch of talk' that meets two criteria: First, each stretch comprises the beginning and the end of an exchange about the meaning of a referent across the turns. Second, it includes a pair of similar gestures produced by different speakers to depict the same referent across the turns. For each pair of similar gestures in a stretch of talk, the judgment on whether one repeated the other rest on 'form' and 'meaning': The high similarity between two gestures in this study was determined by the congruence rates across five gesture features: 'handedness', 'position', 'orientation', 'hand shape', and 'motion' (McNeill 1992, 2005); a replicated gesture also represents the same referent depicted by its corresponding gesture in the prior context.

Only representational gestures conveying substantive meaning were considered in the present study. Example (1) is about the strange behavior of a friend. The stretch of talk has to do with a discussion of the meaning of the ideophone which sounds like [yuyu]. Within the stretch of talk encompassing three turns, a pair of similar gestures is produced by the two participants. In Line 1, the first speaker asserts that the friend behaves in a 'yuyu' way and the ideophone is also depicted in gesture (Fig. 1a): Hands at each side of the body, with fingers together and curled into fists, move slightly up and down alternately. In the next turn in Line 2, the second speaker provides her understanding of the 'yuyu' behavior and grounds it by the use of the new lexical verb *tiàowǔ* 'dance' and the repetition of the first speaker's gesture (Fig. 1b). The third turn in Line 3 is the end of the discussion, as the addressee displays the acceptance of the second speaker's understanding that dancing is the strange behavior of the friend by the use of the agreement marker *duì* 'right' three times.

(1) 1 1st: .. xiànzài biànchéng **yuyu** zhèyiàngzhi
now become IDEOPHONE like this
'Now, (he) became 'yuyu', like this.'
2 2nd: ..ránhoù hái hùi **tiàowǔ**
then also will dance
'(He) will also dance.'
3 1st: .. **duì duì duì**
right right right
'Right, right, right.'

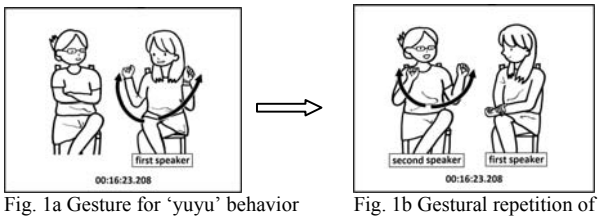


Fig. 1a Gesture for 'yuyu' behavior Fig. 1b Gestural repetition of 1a

Two coders worked separately to identify stretches of talk, each comprising the whole discussion of the meaning of a referent. A total of 63 stretches were found in the data. Among the 24 stretches involving the occurrence of speech-accompanying gestures, there were 12 instances of gestural repetition, constituting twelve co-referential gesture pairs for the study: The initial representational gesture was produced by the 'first speaker'. Later, the 'second speaker' repeated it his/her turn. The similarity between the two instances in each pair was rated on a five-point scale of agreement: The realization of a feature in both gestures of each pair was coded as 'Alike' if the coder chose 'agree' or 'strongly agree', as 'Not Alike' if the judgment was 'disagree' or 'strongly disagree', and as 'Neutral' for the choice 'neutral' on the scale. The judgment was subject to the relatively objective spatio-physical manifestation of the features, and coders reached total agreement on their analysis. See the high congruence rates in Table 1.

Not only does gestural repetition in conversation reveal the interactive nature of grounding, it also manifests what information about the addressee's state of knowledge the second speaker would take into account during the construction of utterance.

Table 1. The congruence rates across the five gesture features

	Hand shape		Handedness		Position	
Alike	12	100%	10	83%	11	92%
Not alike	0	0%	2	17%	1	8%
Neutral	0	0%	0	0%	0	0%
	Motion		Orientation			
Alike	11	92%	8	67%		
Not alike	1	8%	4	33%		
Neutral	0	0%	0	0%		

*('Alike': coders (strongly) agreed that the realization of the feature was similar in both gestures; 'Not Alike': coders (strongly) disagreed that the realization was similar in both gestures; 'Neutral': similarity or difference not noticeable)

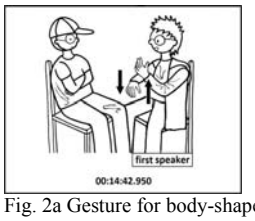
Linguistic-gestural grounding of meaning

How do speakers ground meaning in speech and by gestural repetition? First, it is necessary to consider the stretch of talk encompassing the whole of the exchange which is concerned with the meaning of a referent, and including a pair of similar gestures produced by two different speakers to depict the same referent. On the one hand, the contextual situation that includes the initial occurrence of a replicated gesture, and the fact that the first speaker in the contextual situation typically becomes the addressee when the second speaker performs the act of grounding help understand the situation in which the need for the grounding of meaning arises and the incorporation of linguistic-gestural contextual elements in the second speaker's turn. On the other hand, the consideration of the first speaker's act and of the addressee's later acceptance of the second speaker's idea about the referent are relevant to the discussion of the bilateral process of speaking.

This section presents four kinds of contextual situation, as manifested in the turn of the first speaker, that includes a gesture and that would prompt another participant to provide and ground meaning for a referent having been brought up in the first speaker's turn. It will be shown that despite the situational difference, a particular cross-modal grounding strategy can be found in the second speaker's turn. The talk about the meaning of a referent ends when the addressee displays his/her acknowledgement of the newly established meaning.

In the first kind of situation, when new referents are expressed by demonstratives, non-conventional ideophones, or homonyms, the lexical meaning lacks clarity. In the data, during the discussion about the shape of the body of a friend, a speaker mentions that the girl was fat at the time that he saw her in a college-preparatory cram school, but that she was very thin at the time when he next saw her in the school library in college. At the time that the first speaker tries to provide a reason for the change in the girl's shape, as shown in Example (2), his utterance includes the demonstrative *nàge* 'that' as the main predicate (Line 1) and he simultaneously produces a gesture (Fig. 2a) to specify the meaning: The right hand first rises up with the fingers open from the thigh to the front of the chest; then the left hand starts rising up with the fingers open. Both hands move up and down alternately four

times, depicting the idea of the changes in the body shape. The absence of an explicit lexical meaning for the new gestural referent in the prior context prompts the second speaker to provide an explicit meaning for the demonstrative and ground it by the use of cross-modal resources. The second speaker produces a new lexical verb *tiáojié* 'adjust'; in gesture, he repeats the first speaker's gesture of body-shape-adjustment (Fig. 2b) by copying the shape of hands being in front of the chest, its handedness (both hands) and the direction of movement (up and down). Finally, the addressee, who is also the first speaker in the first turn, displays the acceptance of the newly established meaning by the use of the agreement marker *duì* 'right' in the next turn (Line 3). More elaboration follows the agreement marker in this acceptance phase, providing further information about the idea of adjustment as being a process of becoming fat and then thin and then fat again and thin again over time. The elaboration indicates further confirmation of the newly established meaning and mutual understanding between the participants.



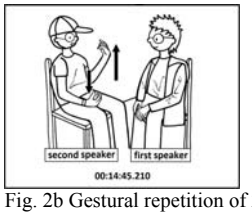
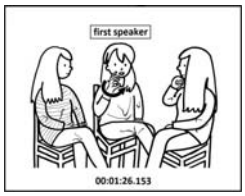


Fig. 2a Gesture for body-shape adjustment Fig. 2b Gestural repetition of 2a

(2) 1 1[s]: ..qíshí tā běnlái jìu mán huì **nàge** a
in fact 3SG originally then quite can that PRT
'Actually she was originally quite good at that.'
2 2[nd]: ..hái mán néng **tiáojié** jìu duì le
still quite can adjust then right PRT
'(She) is quite good at adjusting (her body shape).'
3 1[s]: ..**duì** a tā ..tā yě pàng shòu pàng shòu
right PRT 3SG 3SG also fat thin fat thin
.. zhèyàng
like this
'Right. She…she's fat, and then thin, and then fat, and then thin, and then fat, and then thin, like this.'

In another situation, the first speaker encounters difficulty in verbalization, and the turn-construction unit is incomplete and the meaning is not fully conveyed. Thus, the need arises for the next speaker to provide meaning. Example (3) is about idealization. The first speaker says that if a person were someone with whom she had failed to establish a close relationship with, s/he would idealize the person. A gesture is produced to depict the idea of *lǐxiǎnghuà* 'idealization' in Line 1 (Fig. 3a): The right hand rises to cheek level with fingers slightly apart and bent, after which the hand turns around clockwise. After the assertion, the speaker attempts to further explicate her understanding of 'idealization' yet fails to finish her expression of her thought after her utterance of the second degree adverb *hěn* 'very' (Line 1). To deal with the first speaker's incomplete verbalization, the second speaker provides a meaning for the idea of idealization in the next turn. She presents her own understanding of what may be meant by the idealization of a person by formulating a new statement with a different verb *měihuà* 'beautification' in Line 2. She also grounds her interpretation by gestural repetition, depicting *lǐxiǎnghuà* in a similar way (Fig. 3b). At the end of the discussion, the addressee, who is also the first speaker, does not signal acceptance verbally or gesturally; instead, the talk moves on without opposition. The absence of opposition typically implicates agreement and mutual understanding, and the act of grounding the meaning is still accomplished.



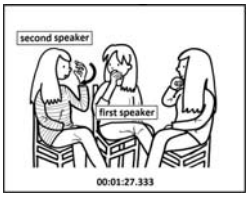


Fig. 3a Gesture for idealization Fig. 3b Gestural repetition of 3a

(3) 1 1[s]: ..nǐ débúdào de dōngxī.. ránhoù yìzhí
2SG NEG.get DE thing then continuously
dùi tā hěn **lǐxiǎnghuà** de.. hěn
to 3SG very idealize PRT very
'For the things that you can't get, then (you) very much keep idealizing him, very'
2 2[nd]: ... nǐ bǎ tā **měihuà**
2SG BA him beautify
'You beautify him.'

The third situation has to do with alignment, in that what the first speaker talks about is recognized by the second speaker. See Example 4.

Fig. 4a Gesture for idealization Fig. 4b Gestural repetition of 4a

(4) 1 1[st]: .. yīnwèi tā yoǔ [**máng** a]
because 3SG have prickle PRT
'Because it has prickles.'
2 2[nd]: ..shàngmiàn yoǔ nàge **háomáo**
on there.be that fine hair
'There are fine hairs on (it).'

In Example (4), the conversational topic is about feeling itchy during the harvesting of rice crops in a field. The first speaker is explaining that there are small thorns or prickles *máng* on the stems in Line 1. At the same time, a gesture is produced for prickles (Fig. 4a): The right hand at shoulder level curves into the palm; the left-hand fingers at chest level come together. The configuration as a whole enacts the holding the stem of a crop on which there are prickles. The second speaker then demonstrates his agreement in Line 2. Again, cross-modal grounding is in evidence: Linguistically, the second speaker provides more semantic information about the crops - having *háomáo* 'fine hairs' on the stems;

manually, the initial gesture produced in the addressee's previous turn is replicated (Fig. 4b), with a repetition of the hand shape of holding a stem with both hands in front of the chest, as well as the orientation of the hands and fingers being together and curved in. As in Example (3), in the acceptance phase, the stretch of talk for the discussion of *máng* and *háomáo* is finished with no opposition from the addressee.

In contrast to the case in the agreement situation, the second speaker may engage in the discussion of a referent because s/he does not agree with what had been uttered by the first speaker in the prior turn. A talk in the data is about the kind of musical instrument played by a character in a movie. As shown in Example (5), the first speaker uses a general term *yuèqí* 'musical instrument' in speech (Line 1) but gestures a particular kind that requires the use of a bow to play (Fig. 5a): The right hand goes up to shoulder level with the fingers curled into a fist as if holding a bow; the left hand rises to waist level, also with fingers curled into a fist as if holding the lower part of the instrument. Then, the right hand moves horizontally to the left one time to enact the playing of a stringed musical instrument that requires the use of a bow. Since the second speaker holds a contrary opinion about the referent *yuèqí* in regard to the instrument played in the movie, he brings up a different understanding and grounds it by gestural repetition along with speech in his turn. In Line 2, he mentions that it is the type that is played with the fingers, as represented in speech by *yuèqìn* 'plucked lute with a wooden body'. But, what is of note is that in gesture, instead of enacting *yuèqìn* which is played with the fingers, the speaker repeats the first speaker's gesture (Fig. 5b), as his right-hand fingers form a fist as for holding a bow, and the right hand moves in the same leftward direction to enact the idea of playing a musical instrument with a bow. After that, the addressee also conveys acceptance by use of the commonly used particle *duì* 'right', followed by more elaboration to indicate re-assurance that what the second speaker has mentioned about the musical instrument is correct.

Fig. 5a Gesture for musical instrument played with a bow

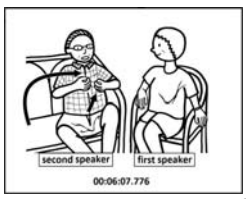


Fig. 5b Gestural repetition of 5a

(5) 1 1st: .. màobó shì nà zǒng... zhūanyiè de nà
Maobo COP that kind professional DE that
zǒng... **yùeqì** de.. nǐ zhīdào ma.. sǔoyǐ
kind musical instrument DE 2SG know PRT so
'Maobo (used) that kind of professional musical instrument, you know. So,'

2 2nd: (0) méiyiǔ.. tā shì nàge...(1.1)jù **xúanzǎi**
NEG 3SG COP that play plucked lute with a wooden body
'No, he played a kind of plucked lute with a wooden body.'

3 1st: .. **duì** a **duì** a... jiu shì nà zǒng
right PRT right PRT EMP COP that kind
...yùeqì de
musical instrument PRT
'Righ, right. It's that kind of musical instrument.'

To summarize, in the four types of contextual situation mentioned above, no matter whether the first speakers' utterances were complete or not, or clearly or vaguely expressed, each utterance was an assertion or assessment about the quality of states, activities, or processes or about the characteristics of people or objects. Each of the utterances included a referent in gesture. Among the total of eleven gestures, six depict concrete entities or activities, such as the gesture for musical instruments, and five enact abstract ideas, such as 'idealization'. All of the gestural referents carry new information, as the corresponding lexical referents had not been brought up previously in the conversation prior to the point at which they were uttered.

In the next turn, when the other participant wants to express the meaning of a referent brought up in the first speaker's utterance, be it gestured or not, be the action prompted by the first speaker's lack of clarity or by the presence of an obstacle to verbalization in the utterance, or by the second speaker's alignment or disagreement with what the first speaker has uttered, "the speaker must design what she says against the current common ground with her partner. His beliefs about their common ground should be coordinated with hers if they are to understand one another efficiently" (Wilkes-Gibbs & Clark, 1992: 184). The way to make such a design in face-to-face conversational interaction is cross-modal, and the new interpretation is provided and grounded through the collaboration between speech and gestural repetition: During the construction of utterance, the second speaker repeats the first speaker's gesture and expresses new meaning for the referent under discussion in speech.

The process of the grounding of meaning could take numerous turns across speakers. When the addressee finally expresses acceptance of a new meaning, the act is accomplished and mutual understanding is reached. Agreement markers such as *duì* 'right' or *o* 'I see' are frequently used to convey agreement with the newly-established meaning. Acceptance can also be signaled non-verbally, for instance by head nods. More elaboration about the referent after the agreement marker further confirms the understanding of the new meaning provided by the second speaker. Finally, the talk could move on to another subject matter without the addressee's overt acceptance. The absence of opposition usually suggests agreement and mutual understanding.

Cross-modal grounding strategies

There are various ways speakers can incorporate gestural signals in the process of grounding. According to Clark and Krych (2004: 79), participants in task-based communication relied on gestural signals like "exhibiting, poising, pointing at, and placing physical objects, nodding and shaking heads, and

directing eye gaze, and on other mutually visible events." As to the participants in face-to-face conversations, when they engage in the act of the grounding of meaning for a referent under discussion, the linguistic-gestural manifestations demonstrate the use of a particular cross-modal strategy: The first speaker's gesture, and sometimes together with the co-referential lexical expression, is replicated to form a semantic foundation shared by the participants; upon such gestural common ground, the speaker readily conveys a new meaning with a new lexical expression. What is of note is that the second speaker would rather repeat the addressee's previous gesture than produce a different one for the new constituent, even though the speaker opposes the meaning or referent being uttered and/or depicted.

Such pairing of the initial gesture and its replicated counterpart and their collaboration with speech to ground meaning in speech communication bears out the *principle of least joint effort* (Clark, 1996; Clark & Brennan, 1991; Clark & Schaefer, 1987; Clark & Wilkes-Gibbs, 1986), in that "[f]ace to face, [people] should exploit that combination of vocal and gestural actions they judge will take the least joint effort" (Clark & Krych, 2004: 64). In this study, the combination of gestural repetition and speech would be the least joint effort the two modalities undertake for the realization of the expression of common ground in gesture and new meaning in speech within a clausal utterance in the second speaker's turn. The cross-modal collaboration would also facilitate the addressee's comprehension of a newly established meaning in the presence of a knowledge foundation already shared between the speaker and the addressee.

Moreover, the use of a grounding strategy based on gestural repetition along with speech supports speaking as a bilateral process to achieve mutual understanding. In a bilateral process, "[s]peakers monitor not just their own actions, but those of their addressees, taking both into account as they speak. Addressees, in turn, try to keep speakers informed of their current state of understanding" (Clark & Krych, 2004: 62). In daily conversation, the second speaker not only provides meaning for the referent at issue, but s/he also grounds the new information by considering the addressee's current state of knowledge at the same time, as evidenced by the occurrence of gestural repetition. In other words, as the first speaker mostly becomes the addressee during the second speaker's construction of utterance, the initial occurrence of a replicated gesture is produced by the addressee in his/her prior turn as the first speaker. Even though the addressee is a third participant, s/he in the same discourse environment would have recognized the gesture produced by the first speaker. Thus, the repetition of the gesture constitutes a semantic foundation of knowledge shared by both the speaker and the addressee, based on which a new meaning of the referent at issue is readily conveyed in speech.

Finally, addressees also play an active role in the bilateral process of speaking, as "[they] take an active part both: (1) by telling speakers about their understanding and (2) by giving them access to evidence of understanding" (Clark & Krych, 2004: 77). In the conversational data, after the second speaker has finished the gestural repetition and presentation of new meaning, the addressee displays understanding of the new meaning. If the addressee shows lack of understanding, then the grounding of meaning continues and more turns are required. The acceptance phase comes when confirmation of understanding is expressed. The use of agreement markers and head nods is common. In addition, more elaboration can be provided to further indicate acceptance. For the talk to go on in spite of the lack of the addressee's explicit display of acceptance of the new meaning implicates agreement with no opposition. All of the various types of responses given by the addressees, again, support the bilateral process of speaking.

Conclusion

In daily conversation, the use of replicated gestures is not frequent. One reason for this is that speakers have numerous communicative goals and perform many actions other than providing the grounding of meaning for mutual understanding. Another reason is that speakers do not necessarily repeat others' gestures at the time they ground semantic information about the same referent. Nevertheless, the use of replicated gestures is by no means a matter of chance. In Holler and Wilkin (2011), the study required the two participants to focus their talk on referents, "in order to figure out whether they are talking about the same thing" (ibid, 136). A total of 113 occurrences of gestural repetition were found. While replicated gestures in speech communication do not occur by chance, the present study has shown that their occurrence and collaboration with speech indicates the use of a cross-modal strategy for the grounding of meaning. The use of such grounding strategy, in turn, bears out the bilateral nature of speaking in that it manifests that the speaker takes into account the addressee's knowledge state to form a semantic common ground across the turns during the construction of their own utterance. The addressee in the next turn also informs the speaker about his/her understanding. Whether this cross-modal grounding strategy used by Mandarin speakers in conversation is language-specific or not awaits future studies across different languages.

Acknowledgments

This research was funded by grants from the National Science Council, Taiwan (NSC 97-2410-H-004-111-MY3).

Referents

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language* , *501*, 62-81.

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S.D. Teasley (Eds.), *Perspectives on Socially Shared Cognition*

(pp. 127–149). Washington, DC: American Psychological Association.
Clark, H. H., & Schaefer, E. F. (1987). Concealing one's meaning from overhearers. *Journal of Memory and Language*, *262*, 209–225.
Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science: A Multidisciplinary Journal*, *132*, 259–294.
Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *221*, 1–39.
de Fornel, M. (1992). The return gesture: Some remarks on context, inference, and iconic gesture. In P. Auer, & A. Di Luzio (Eds.), *The Contexualisation of Language* (pp. 159-176). Amsterdam: John Benjamins,
Holler, J., & Wilkin, K. (2011). Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *Journal of Nonverbal Behaviour*, *35*, 133–153.
Kendon, A. (2004). *Gesture: Visible Action as Utterance.* Cambridge: Cambridge University Press.
Kimbara, I. (2006). On gestural mimicry. *Gesture*, *61*, 39–61.
Koschmann, T., & LeBaron, C. (2002). Learner articulation as interactional achievement: Studying the conversation of gesture. *Cognition Instruction*, *20*, 249-282.
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.
Parrill, F., & Kimbara, I. (2006). Seeing and hearing double: The influence of mimicry in speech and gesture and observers. *Journal of Nonverbal Behaviour*, *304*, 157–166.
Tabensky, A. (2001). Gesture and speech rephrasings in conversation. *Gesture*, *12*, 213–235.
Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Cognition*, *31*, 183-194.