

Moving past the lexical affiliate with a frame-based analysis of gesture meaning

Schuyler Laparle¹

¹Tilburg University

s.m.laparle@tilburguniversity.edu

Abstract

Interpreting the meaning of co-speech gesture often involves identifying a gesture's 'lexical affiliate', the word or phrase to which it most closely relates (Schegloff 1984). Though there is work within gesture studies that resists this simplex mapping of meaning from speech to gesture (e.g. de Ruiter 2000; Kendon 2014; Parrill 2008), including an evolving body of literature on recurrent gesture and gesture families (e.g. Fricke et al. 2014; Müller 2017), it is still the lexical affiliate model that is most apparent in formal linguistic models of multimodal meaning (e.g. Alahverdzhieva et al. 2017; Lascarides and Stone 2009; Pustejovsky and Krishnaswamy 2021; Schlenker 2020). In this work, I argue that the lexical affiliate should be carefully reconsidered in the further development of such models.

In place of the lexical affiliate, I suggest a further shift toward a frame-based, action schematic approach to gestural meaning in line with that proposed in, for example, Parrill and Sweetser (2004) and Müller (2017). To demonstrate the utility of this approach I present three types of compositional gesture sequences which I call *spatial contrast*, *spatial embedding*, and *cooperative abstract deixis*. All three rely on gestural context, rather than gesture-speech alignment, to convey interactive (i.e. pragmatic) meaning. The centrality of gestural context to gesture meaning in these examples demonstrates the necessity of developing a model of gestural meaning independent of its integration with speech.

Index Terms: discourse, pragmatics, interactive gesture

1. Introduction

The lexical affiliate model of co-speech gesture meaning presupposes that the meaning of a gesture can be deduced from the meaning conveyed in speech, so long as one can identify a co-expressive word or phrase.¹ For example, if a speaker moves their hand upward as they say "we climbed up the tower" or "rent prices are going up", it seems obvious that the upward movement of the hand iconically represents the directional concept that is co-expressed by the lexical item "up", concretely in the first example and metaphorically in the second.² As such, the lexical affiliate model would suggest an interpretation of the gesture as also conveying the directional concept UP.

Two observations are central to the appeal of this model. First is the temporal alignment of a gesture with its lexical affiliate – even when a gesture slightly precedes its proposed affiliate, some portion overlaps (e.g. Harrison 2010; McNeill 1985, 2005; Schegloff 1984). Usually, this overlap is during the gesture's 'stroke', often considered the most meaningful aspect of the gesture's articulation (Kendon 1972). Second is the intu-

ition that the meaning of gesture is less pre-specified and more subject to idiosyncratic variation than speech (McNeill 2005; Parrill 2008). These two observations drive the lexical affiliate model – because of the consistent temporal alignment and because of the difficulty of interpreting gesture without first interpreting speech, the meaning of a gesture must be *dependent* on the meaning of speech.

My goal in this work is to highlight ways in which both of the above claims are detrimental to the development of formal multimodal theories of language. Instead of assuming the centrality of spoken meaning in a multimodal message, I suggest that gesture meaning can, and should, be analyzed independently, prior to analyzing its alignment or integration with speech. The model I propose in this work is in line with action schematic approaches to gesture meaning (e.g. Cienki 2013; Mittelberg 2018; Müller 2017), as well metaphoric approaches to contextualized gesture interpretation (e.g. Bressemer and Müller 2017; Bressemer and Wegener 2021; Ladewig 2011).

The remainder of this paper is structured as follows. First, I discuss shortcomings of the lexical affiliate model in more detail, focusing on *misalignments* between gesture and speech. I then outline an action schematic frame-based analysis of gesture meaning that can be used to interpret gesture on its own terms, prior to gesture-speech integration. In support of this analysis, I present three case studies from the American talk show *The Late Show with Stephen Colbert* in which the interpretation of interactive gestures relies more on the consideration of gestural context than gesture-speech alignment. I present these case studies as three types of compositional meaning in gesture which I call *spatial contrast*, *spatial embedding*, and *cooperative abstract deixis*.

2. The lexical affiliate

As described above, the notion of a 'lexical affiliate' is often used when interpreting gestures co-occurring with speech, especially in discussions of semantic meaning. It seems to me that this approach has been favored because it (i) allows us to use pre-existing models of meaning, and (ii) is consistent with gesture-speech alignment. Because we have developed sound theoretical models of spoken and written meaning, it is sensible to try to expand these models to also accommodate gestural meaning. The temporal alignment of gesture to speech then serves to legitimize this decision. However, these conveniences begin to break down if one takes the independent semiotic capacity of gesture more seriously (e.g. Kendon 2014).

The two empirical shortcomings I am concerned with in this work are cases of *semiotic enrichment* and *discourse management*. In the first, the gesture expresses meaning beyond that conveyed by its lexical affiliate. In the second, the gesture frequently lacks a clear lexical affiliate entirely. I discuss each of these in turn (Sections 2.1 & 2.2). The latent predominance of the lexical affiliate model also has a theoretical shortcoming – relying on speech to interpret gesture perpetuates a concep-

¹This excludes 'emblems', gestures that convey pre-specified conventionalized meaning and that are not dependent on co-occurring speech for interpretation (Kendon 2004; McNeill 2005).

²See McNeill (2005) for discussion of iconic gestures as 'concrete' or 'metaphoric'.

tion of language-as-speech, equating linguistic expression with speech and writing *a priori*. The negative impacts of this false equivalency are discussed in Section 2.3.

2.1. Semiotic enrichment and gesture

When discussing gestures that convey primarily semantic information (i.e. representational gestures), a lexical affiliate is often readily identifiable. For instance, in the example given in the introduction, it seems in some way obvious that an upward movement of the hand co-expresses the meaning conveyed by a word like “up”. However, the semantic overlap is rarely perfect. Frequently, the gesture provides information beyond that expressed by its lexical affiliate (e.g. de Ruiter 2000; Kendon 2014). I refer to this additional meaning as *semiotic enrichment*.

Gesture is particularly well-suited for expressing spatial information and aspects of physical action such as manner (e.g. Hockett 1978). Because of this efficiency, gestures that depict action and spatial arrangements are rarely fully redundant with the speech they accompany. Kendon (2014) provides a particularly apt example of this in which two radically different gestures are both affiliated with the word “throw”. In the first, the gesturer repeatedly flicks a loosely-closed hand to the side at waist level, iconically enacting the throwing of rice. In the second, the gesturer moves a cupped hand from waist level up over his shoulder, as if to lob an object backward, iconically enacting the throwing of oranges. In both cases, the manner and trajectory of the throwing action is not expressed in speech. Instead, we rely on world knowledge to envision different types of throwing events. The manner and trajectory is, however, overtly and efficiently expressed by the accompanying gesture.

In cases of semiotic enrichment, such as those described above, maintaining that a gesture’s meaning can be fully deduced from a lexical affiliate is difficult. Even when the gesture is *partially* redundant with the accompanying speech, there is still gestural meaning that cannot be reduced to that conveyed in the spoken mode. In order to interpret this additional meaning, we must then consider gesture meaning *independently*. If this is possible, and indeed necessary, then we have reason to pursue full independent models of gestural meaning.

2.2. Interactive gesture and discourse management

The insufficiency of the lexical affiliate approach to gestural meaning is especially apparent when considering *interactive gestures*, gestures that primarily convey pragmatic meaning and help to “maintain the conversation as a social system” (Bavelas et al. 1992:469). Recognized functions of interactive gestures include signaling turn structure (e.g. Streeck and Hartge 1992), speaker attitude (e.g. Wehling 2017), and coherence relations (e.g. Hinnell 2019; Jannedy and Mendoza-Denton 2005). Interactive gestures are pervasive and contribute to discourse management at every level of discourse and interactional structure.

Despite their frequency, interactive gestures are relatively understudied and often perceived of as “unruly” in their apparent idiosyncrasy (Streeck 2009:181). This perception is partially driven by the inability to rely on lexical affiliates in the interpretation of interactive gestures (e.g. Laparle 2022). Very often, interactive gestures appear to lack a lexical affiliate completely. When a lexical affiliate can be identified, it is most often a multifunctional discourse marker like “so” or “but”. As such, if one searches for the meaning of a gesture in the speech it accompanies, interactive gestures appear to misbehave.

Consider, for example, a gesture in which an upturned open hand is stretched toward an addressee, as if to present an ob-

ject for inspection. This gesture, frequently referred to as the ‘palm-up open hand’ (PUOH) gesture, is cross-linguistically common and highly multifunctional (Cooperrider et al. 2018; Müller 2004). The PUOH gesture can express, for example, the offering of new information, a request for new information, or an acknowledgement of a previous contribution. One can imagine lexical expressions for each of these functions, such as “here’s an idea”, “what’s your opinion”, and “as you said earlier”. However, in these cases, the connection between the meaning of the gesture and the meaning of the phrase is unclear.

With representational (i.e. semantic) gestures, the connection between the lexical affiliate and the gesture is usually obvious – the throwing gestures are affiliated with the word “throw” because they iconically resemble the concept expressed by the word “throw”. The PUOH gesture does not resemble the concept of *providing information*, *asking for an opinion* or *citing*, and yet one can sense an association. This difficulty in identifying and interpreting the lexical affiliate of interactive gestures is exactly what makes interactive gestures particularly helpful for developing a model of gestural meaning independent of gesture-speech alignment. The analysis of interactive gesture encourages the development of a new model of gestural meaning because the old model simply does not work.

2.3. Avoiding theoretical bias

As a final critique of the lexical affiliate approach, I will briefly discuss the ways in which the lexical affiliate reinforces problematic conceptions of language and what counts as ‘linguistic’.

Dominant models of meaning in language were built with spoken and written language in mind.³ Kinesic modes, such as sign and gesture, were not seriously considered, thus instilling an inherent bias toward equating ‘linguistic’ meaning with spoken and written meaning. This bias contributed directly to the conception of sign languages as less than ‘full’ languages; signed meaning is *different* from spoken and written meaning in its use of iconicity, and thus ‘non-linguistic’ given existing models (Wilcox and Occhino 2017). Though sign languages are now perceived of as full languages by the linguistic community, dominant models of meaning have remained largely unchanged.

Gestures, like signs, take advantage of iconicity in ways that are not possible in speech and writing (e.g. Hockett 1978). Arguing about the linguistic status of gestural meaning, with or without accompanying speech, thus runs into the same issue that sign language linguistics has had to contend with – iconic meaning is ‘non-linguistic’ because it is not the type of meaning that existing models are built for. When we analyze gestural meaning via a lexical affiliate, we reinforce the conception of iconic meaning as ‘non-linguistic’ by requiring gestures to mean via a spoken unit rather than via their own iconicity.

One might argue that I am being too narrow in my understanding of the lexical affiliate; perhaps it serves more as a ‘guideline’ than as a ‘model’ of gestural meaning. Indeed, I have not seen work explicitly arguing *for* a lexical affiliate model. However, this is exactly the problem. The notion of the lexical affiliate has been used for so long that it has become a kind of assumption about meaning in gesture. The lexical affiliate as *assumption* is most clearly reflected in many recent attempts to model gesture speech integration – the meaning of the gesture is not considered independently. Instead, each gesture is mapped to a spoken unit, perhaps augmenting the spoken unit with additional spatial meaning, but still secondary to speech.

³See Linell (2005) for potent discussion of the ‘written language bias’ in linguistic theory.

The ‘embodied turn’ in linguistics (Neville 2015) recognizes language as a fundamentally multimodal system and provides the field with an opportunity to scrutinize long-held conceptions of what language and ‘linguistic’ meaning is (e.g. Ferrara and Hodge 2018; Gabarró-López and Meurant 2022; Perniss 2018). It has also been repeatedly pointed out that this turn toward embodiment and multimodality provides a unique opportunity to strengthen our existing theories by restructuring them to accommodate new types of data and meaning (e.g. Cienki 2022; Mondada 2016; Schoonjans 2017; Sweetser 2007). To take advantage of this opportunity, I believe that it is necessary to model gestural meaning by foregrounding the unique iconic capacities of gesture rather than gesture-speech alignment.

3. An alternative to the lexical affiliate

This section presents a way to move beyond the lexical affiliate by modelling gestural meaning based on the particular semiotic capacities of gesture itself. The model I propose is built on a *frame semantic* approach to meaning (Fillmore 1976) and an *action schematic* approach to gesture analysis (Cienki 2013; Mittelberg 2018; Müller 2017). For subsequent integration with the spoken mode, I argue for a conceptual integration approach reminiscent of that proposed in Parrill and Sweetser (2004).

3.1. Gesture meaning as action schematic frames

Frame semantics, as outlined in Fillmore (1976), proposes that linguistic elements convey meaning by evoking experiential frames rather than by activating a conventionalized dictionary-like definition. Under this framework, a word like “throw” does not mean “to propel through the air by a forward motion of the hand and arm” (Merriam-Webster), per say, nor does it mean via a logic-based representation. Instead, the word “throw” means via generalizations we have made over our embodied experiences of throwing events. This meaning is formally represented as a list of *frame elements*, the participants, entities, states, and events that are associated with these experiences. The frame evoked by “throw”, represented in small caps as THROW, thus consists of a thrower, a thrown object, the action of moving one’s arm, and the resulting movement of the object along an arcing trajectory. The THROW frame can be considered an action schematic frame in that it represents the participants, processes and results of a particular, physically grounded, action.⁴

A gesture evokes an action schematic frame through *iconicity*, i.e. by physically resembling some aspect of the action (e.g. Müller 2017). Consider again Kendon’s (2014) ‘throwing’ gestures. Under an action schematic analysis, these gestures mean ‘throw’ because they physically resemble throwing events, not because they align with the lexical item “throw”. Given a frame semantic approach, both the gestures and the lexical item “throw” evoke the THROW frame, but they do so independently. This allows the throwing gestures in Kendon’s examples, but *not* the word “throw”, to evoke more highly articulated frames in which the manner and trajectory of the throwing action is specified.

In the present work, I am concerned with two action schematic frames, PRESENTATION and ATTENTION DIRECTION, both of which are evoked frequently by interactive gestures. The PRESENTATION frame is evoked by gestures that resemble the presentation of a physical object for inspection. The

PUOH gesture that was previously discussed is one such gesture. The ATTENTION DIRECTION frame is evoked by deictic gestures, such as pointing, which serve to orient mutual attention toward a single region of space. The frame elements of each are listed in Table 1.

PRESENTATION	ATTENTION DIRECTION
Object(x)	Object(x)
Participant: presenter	Participant: director
Participant: observer	Participant: observer
Location(y)	Location(y)
Action: present object(x) into location(y)	Action: direct attention to object(x) in location(y)

Table 1: Action schematic frames

As with Kendon’s throwing gestures, particular formal features of a presentational or deictic gesture can evoke a more elaborated PRESENTATION or ATTENTION DIRECTION frame than those given in Table 1. For example, a PUOH gesture (Müller 2004) and a pinching gesture (Kendon 1995) would both evoke the PRESENTATION frame because both of their physical attributes are consistent with the presentation of an object into a particular region of space. However, the features of the presented objects are distinct. A PUOH gesture is consistent with the presentation of a relatively light, medium-sized object that is otherwise underspecified for physical features. A pinching gesture, on the other hand, is consistent with the presentation of a small delicate object that requires close proximity of the observer for proper inspection.

3.2. Contextualized meaning and metaphor

Action schematic frames are well-suited for modelling gesture meaning based on its iconic semiotic capacities. However, action schematic frames alone do not fully account for how a gesture is used in context to convey semantic meaning or perform discourse management. Following Parrill and Sweetser (2004), I propose that the communicative *function* of a gesture is interpreted in context by identifying a ‘target frame’ that profiles the communicative goals in a particular moment of an interaction. This can be represented as blending (Fauconnier and Turner 2008) or conceptual metaphor mappings (Lakoff and Johnson 1980). I opt for the second here for the sake of simplicity.

Table 2 shows the process of deriving a contextualized function of a presentational gesture. The presentational gesture itself conveys meaning by evoking the PRESENTATION frame. This serves as the ‘source frame’. The intended communicative message is then derived by relating the PRESENTATION frame to that of the intended communicative process (e.g. TOPIC INTRODUCTION) which serves as the target frame. Equivalencies, called ‘mappings’, are then made between frame elements in each frame. Mappings are represented as leftward arrows.

TOPIC INTRODUCTION		PRESENTATION
Topic(x)	←	Object(x)
Participant: contributor	←	Participant: presenter
Participant: addressee	←	Participant: observer
Discourse status(y)	←	Location(y)

Table 2: Metaphoric mappings

Under this analysis, a presentational gesture performed in a discourse context in which a new topic is introduced is understood as metaphorically presenting the topic as an object for

⁴Though semantic frames may profile non-action-oriented concepts, such as entities and attributes, it is action schematic frames that are most relevant for developing an independent model of gestural meaning.

inspection. The speaker who introduces the topic is understood as the presenter of the object, the topic is understood as the presented object, and the discourse status of the topic (i.e. that it is under discussion) is understood as the physical location of the metaphoric object (i.e. central to both participants).

The same process occurs for interpreting a spoken word or phrase in context. For example, a speaker may begin an explanation by saying “to give you an example”. The lexical item “give” evokes an action schematic frame of OBJECT TRANSFER. Given the lack of literal object transfer, this frame is interpreted metaphorically. The action schematic frame is mapped to a communicative frame in which the “example” is metaphorically understood as the transferred object, and the addressee is metaphorically understood as a recipient.

The important point here is that the meaning of gesture and the meaning of speech align not because of a hierarchical relationship, but because they share the same communicative context (i.e. are mapped to the same target frame) and contribute to achieving the same communicative goal. Under the present analysis, a shared communicative context is understood as a shared target frame. Lexical affiliates of semantic gestures seem ‘obvious’ because the gesture and word evoke the same source and target frames. Interactive gestures resist lexical affiliates even when there is a phrase that performs the same communicative function because the phrase does not typically evoke the action schematic source frame evoked by the gesture.

4. Compositional gesture sequences

In this section, I demonstrate the utility of a frame-based analysis of gesture meaning by considering gesture sequences that evoke a chain of related action schematic frames. These sequences cannot be fully understood by considering their association with the accompanying speech signal. Instead, the gestures mean by evoking a complex, multi-unit action schema.

The first example shows an action sequence expressing *spatial contrast* in which metaphoric objects are presented into different regions of space. The second demonstrates *spatial embedding* in which increasingly specific topics are referred to as objects in increasingly specific regions of space. The third demonstrates *cooperative abstract deixis* in which a topic is referred to as a metaphoric object in space by both interlocutors.

4.1. Spatial contrast

Contrasting two topics or referents is recurrently expressed in gesture via a change in gesture position and bodily orientation (e.g. Calbris 2008; Hinnell 2019). When CONTRAST is expressed by hand gestures, it is typically through deictic gestures, such that attention is directed to two distinct regions of space, or presentational gestures, such that metaphoric objects are introduced into distinct regions of space.

In the following example, American musician LL Cool J produces two presentational gestures into different regions of space, shifting his presenting hand from his left to his right, and shifting the orientation of his body and gaze. These two gestures are depicted in the first two screenshots in Figure 1 (G1 & G2).



Figure 1: Gesture sequence corresponding to ex. 1

As he performs this gesture sequence, he describes two alternatives in speech (“make a lot of money” and “be really famous”), marked as contrastive by an “either...or” construction. The alignment of the gestures depicted in Figure 1 with the speech in this discourse segment is provided in (1). The beginning and end of each gesture is marked by asterisks (“*”). The ‘gesture tier’ below each relevant line of transcript is co-indexed with the screenshots in Figure 1.⁵

- (1) TRANSCRIPT 1: LL COOL J
 [UID:960813f4-11b1-11ea-bff4-089e01ba0335,1854]
 1 because I think we on this planet
 2 to maximize our potential. We’re
 3 not just here to just *either
 *G1- - ->
 4 make a lot of money* *or be
 <- - - - - * *CG2->
 5 really famous* *and that’s it*
 <- - - - - * *CG3 - - - - *

Crucially, the second gesture does not express contrast through its alignment with “or”. The contrast is instead expressed via the relative spatial relationship that holds between G1 and G2. The metaphoric object presented by G1 is distinct from the metaphoric object presented in G2 because they are represented as objects occupying non-overlapping regions of space. If he instead repeated a presentational gesture in the *same* region of space, the introduction of two referents would still be expressed in gesture. However, the contrastive relationship between the two referents would no longer be expressed gesturally, even though a presentational gesture would still align with the contrastive signal “or”.

The final gesture in this sequence (G3) evokes a CLEARING action that has been associated with exhaustivity, as if the gesturer is clearing the space of any possibilities that are not specified (Harrison 2010; Kendon 2004). Once again the deictic meaning of this gesture builds on the composition of the LL Cool J’s constructed space. The exhaustivity conveyed is related to the two contrasting referents that have been introduced into the subsequently ‘cleared’ space.

4.2. Spatial embedding

In *spatial embedding*, the gesturer performs consecutive deictic gestures toward a single region of space or sub-regions therein. Gestures that perform spatial embedding can either be *primarily* deictic, such as in index finger points which do not exhibit physical affordances consistent with another action schema, or *secondarily* deictic, such as presentation gestures which evoke a PRESENTATION frame that introduces a metaphoric object into a particular region of space. In the example below, show host Stephen Colbert produces both variants, beginning with a presentation gesture and proceeding to an index finger point.

As Colbert introduces a new main topic, his vacation home, he performs a two-handed presentational gesture, as if to hold up a round medium sized object for inspection (fig. 2, G1). As he specifies an aspect of his vacation home, he drops one hand, deictically indicating a sub-region of the space he has just indexed as his vacation home (G2). When he gets to an important detail in his story, he shifts handshapes to an index-finger point

⁵As pointed out by a reviewer, prosodic structure and gesture are closely related. Prosodic information is not included here, as I intend to focus on the spatial sequences in the gestural mode. Prosody would have to be considered in a full analysis of the communicative context.

7. References

- Alahverdzhieva, K., Lascarides, A., and Flickinger, D. (2017). Aligning speech and co-speech gesture in a constraint-based grammar. *Journal of Language Modelling*, 5.
- Bavelas, J. B., Chovil, N., Lawrie, D. A., and Wade, A. (1992). Interactive gestures. *Discourse processes*, 15(4):469–489.
- Bressem, J. and Müller, C. (2017). The “negative-assessment-construction”—a multimodal pattern based on a recurrent gesture? *Linguistics Vanguard*, 3(s1).
- Bressem, J. and Wegener, C. (2021). Handling talk: A cross-linguistic perspective on discursive functions of gestures in German and Savosavo. *Gesture*, 20(2):219–253.
- Calbris, G. (2008). From left to right...: Coverbal gestures and their symbolic use of space. In *Metaphor and gesture*, pages 27–53. John Benjamins.
- Cienki, A. (2013). Image schemas and mimetic schemas in cognitive linguistics and gesture studies. *Review of Cognitive Linguistics. Published Under the Auspices of the Spanish Cognitive Linguistics Association*, 11(2):417–432.
- Cienki, A. (2022). The study of gesture in cognitive linguistics: How it could inform and inspire other research in cognitive science. *Wiley Interdisciplinary Reviews: Cognitive Science*, page e1623.
- Cooperrider, K., Abner, N., and Goldin-Meadow, S. (2018). The palm-up puzzle: Meanings and origins of a widespread form in gesture and sign. *Frontiers in Communication*, 3:23.
- de Ruiter, J. (2000). The production of gesture and speech. In McNeill, D., editor, *Language and gesture*, pages 248–311. Cambridge University Press.
- Debras, C. (2017). The shrug: Forms and meanings of a compound enactment. *Gesture*, 16(1):1–34.
- Fauconnier, G. and Turner, M. (2008). *The way we think: Conceptual blending and the mind's hidden complexities*. Basic Books.
- Ferrara, L. and Hodge, G. (2018). Language as description, indication, and depiction. *Frontiers in Psychology*, 9:716.
- Fillmore, C. J. (1976). Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the origin and development of language and speech*, volume 280, pages 20–32. New York.
- Fricke, E., Bressem, J., and Müller, C. (2014). Gesture families and gestural fields. In Müller, C., Cienki, A., Fricke, E., Ladewig, S., McNeill, D., and Tessendorf, S., editors, *Body–language–communication: An international handbook on multimodality in human interaction*, volume 2, pages 1630–1640. Berlin/Boston: De Gruyter Mouton.
- Gabarró-López, S. and Meurant, L. (2022). Contrasting signed and spoken languages: Towards a renewed perspective on language. *Languages in Contrast*, 22(2):169–194.
- Harrison, S. (2010). Evidence for node and scope of negation in coverbal gesture. *Gesture*, 10(1):29–51.
- Hinnell, J. (2019). The verbal-kinesic enactment of contrast in North American English. *The American Journal of Semiotics*, 35:55–92.
- Hockett, C. F. (1978). In search of Jove's brow. *American speech*, 53(4):243–313.
- Jannedy, S. and Mendoza-Denton, N. (2005). Structuring information through gesture and intonation. *Interdisciplinary studies on information structure*, 3:199–244.
- Kendon, A. (1972). Some relationships between body motion and speech. In Siegman, A. and Pope, B., editors, *Studies in dyadic communication*, pages 177–210. Elmsford, NY: Pergamon.
- Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of pragmatics*, 23(3):247–279.
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.
- Kendon, A. (2014). Semiotic diversity in utterance production and the concept of ‘language’. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651).
- Ladewig, S. (2011). Putting the cyclic gesture on a cognitive basis. *CogniTextes. Revue de l'Association française de linguistique cognitive*, (Volume 6).
- Lakoff, G. and Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Laparle, S. (2022). *The shape of discourse: how gesture structures conversation*. PhD thesis, University of California, Berkeley.
- Lascarides, A. and Stone, M. (2009). Discourse coherence and gesture interpretation. *Gesture*, 9(2):147–180.
- Linell, P. (2005). *The written language bias in linguistics: Its nature, origins and transformations*. Routledge.
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychological review*, 92(3):350–371.
- McNeill, D. (2003). Pointing and morality in Chicago. In Kita, S., editor, *Pointing: Where language, culture, and cognition meet*, pages 293–306. Psychology Press.
- McNeill, D. (2005). *Gesture and thought*. University of Chicago Press.
- McNeill, D., Cassell, J., and Levy, E. T. (1993). Abstract deixis. *Semiotica*, 95(1-2):5–20.
- Mittelberg, I. (2018). Gestures as image schemas and force gestalten: A dynamic systems approach augmented with motion-capture data analyses. *Cognitive semiotics*, 11(1).
- Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics*, 20(3):336–366.
- Müller, C. (2004). Forms and uses of the palm up open hand: A case of a gesture family. *The semantics and pragmatics of everyday gestures*, 9:233–256.
- Müller, C. (2017). How recurrent gestures mean: Conventionalized contexts-of-use and embodied motivation. *Gesture*, 16(2):277–304.
- Nevile, M. (2015). The embodied turn in research on language and social interaction. *Research on Language and Social Interaction*, 48(2):121–151.
- Parrill, F. (2008). Form, meaning, and convention: A comparison of a metaphoric gesture with an emblem. *Metaphor and gesture*, pages 195–217.
- Parrill, F. and Sweetser, E. (2004). What we mean by meaning: Conceptual integration in gesture analysis and transcription. *Gesture*, 4(2):197–219.
- Perniss, P. (2018). Why we should study multimodal language. *Frontiers in psychology*, 9:1109.
- Pustejovsky, J. and Krishnaswamy, N. (2021). Embodied human computer interaction. *KI-Künstliche Intelligenz*, 35(3-4):307–327.
- Schegloff, E. (1984). On some gesture's relation to talk. In Atkinson, J. M. and Heritage, J., editors, *Structures of social action: Studies in conversation analysis*, pages 266–296. Cambridge University Press.
- Schlenker, P. (2020). Gestural grammar. *Natural Language & Linguistic Theory*, 38(3):887–936.
- Schoonjans, S. (2017). Multimodal construction grammar issues are construction grammar issues. *Linguistics Vanguard*, 3(s1).
- Streeck, J. (2009). *Gesturecraft: The manufacture of meaning*, volume 2. John Benjamins Publishing.
- Streeck, J. and Hartge, U. (1992). Previews: Gestures at the transition place. *The contextualization of language*, pages 135–157.
- Sweetser, E. (2007). Looking at space to study mental spaces co-speech gesture as a crucial data source in cognitive linguistics. *Methods in cognitive linguistics*, 18:201–224.
- Wehling, E. (2017). Discourse management gestures. *Gesture*, 16(2):245–276.
- Wilcox, S. and Occhino, C. (2017). Signed languages. In Dancygier, B., editor, *The Cambridge Handbook of Cognitive Linguistics*, page 99–117. Cambridge University Press.