

John Benjamins Publishing Company



This is a contribution from Gesture 16:2
© 2017. John Benjamins Publishing Company

This electronic file may not be altered in any way.
The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible only to members (students and faculty) of the author's/s' institute. It is not permitted to post this PDF on the internet, or to share it on sites such as Mendeley, ResearchGate, Academia.edu.

Please see our rights policy on <https://benjamins.com/content/customers/rights>

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).
Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Embodied frames and scenes

Body-based metonymy and pragmatic inferencing in gesture

Irene Mittelberg

RWTH Aachen University

This paper lays out the foundations of a frame-based account of gesture pragmatics through detailing how frames and metonymy interact not only in motivating gestural sign formation, but also in guiding crossmodal processes of pragmatic inferencing. It is argued that gestures recruiting frame structures tend to profile deeply embodied, routinized aspects of *scenes* (in the Fillmorean sense of the term), that is, of the motivating context of frames. Two kinds of embodied frame structures situated at different levels of abstraction, schematicity, and entrenchment are proposed: (A) *Basic physical action and object frames* understood as directly experientially grounded frames involving physical action and interaction with the material and social world; (B) *Complex, highly abstract frame structures* that are more detached from the motivating contexts of experience. It is further suggested that gestures exhibiting a comparably low degree of iconicity and/or indexicality are likely to assume pragmatic rather than referential functions. Finally, potential avenues for further research into the relation of scenes, frames, and (multimodal) constructions are outlined.

Keywords: frames, scenes, metonymy, metaphor, pragmatic inferencing, motivation, indexicality, multifunctionality

Introduction

Gestural communicative acts are inherently embodied and indexical. So, in principle, they can only be fully accounted for by viewing them through a pragmatic lens, and thus by considering aspects of multimodal discourse pragmatics and the physical, contextual, and sociocultural anchoring of human cognition and situated meaning-making. This paper examines pragmatic functions of gestures, situating them within the larger context of multimodal interaction. It centers on

experiential and cognitive-pragmatic forces that condition the formation, use, and understanding of gestures spontaneously produced with speech. Coverbal gestures here are understood as discourse-integrated hand (and arm) configurations and movements, as well as head motion and whole-body enactments, that have some communicative function (e.g., Kendon, 2004; Müller, 1998).

More specifically, this paper lays out the foundations of a frame-based account of gesture pragmatics through detailing the roles that scenes (Fillmore, 1977), frames (Fillmore, 1982), and metonymy (e.g., Barcelona, 2009; Panther & Thornburg, 2003) may assume in crossmodal processes of pragmatic inferencing. As emphasized by relevance theorists (e.g., Sperber & Wilson, 1995) and cognitive linguists (e.g., Coulson, 2001; Dancygier & Sweetser, 2014; Fauconnier & Turner, 2002; Gibbs, 1994; Lakoff, 1987; Sweetser, 1990), embodied cognitive processes that are operative in the interpretation of linguistic communicative acts strongly rely on inferential mechanisms, which are guided by frames, metonymy, and metaphor to varying degrees. While there exist various definitions of metonymy in the cognitive linguistic literature (for recent overviews see, e.g., Benczes et al., 2011; Littlemore, 2015), metonymy here is, in a first approach, understood as involving two entities pragmatically correlated within the same frame or experiential domain, of which the linguistically or gesturally profiled element enables inference of the other, targeted element. For instance, through embodied metonymic correlations and chains, a gestural typing action may evoke a virtual contiguous keyboard as well as the ensuing email exchange, that is, the EMAIL frame as a whole. Metaphor, by contrast, is assumed to involve a mapping between two different frames or experiential domains, e.g., in *she easily got a grasp of the concept* the mental process of understanding is understood and linguistically expressed in terms of a physical action (e.g., Lakoff & Johnson, 1980; Sullivan, 2013; Dancygier & Sweetser, 2014). Supporting a multimodal and strongly embodied account of these pervasive mechanisms of conceptualization, understanding, and language use, it will be argued that gestures, be they literal or figurative, tend to be driven by frames and metonymy in one way or another. Evidence for this assertion will be presented throughout the paper, putting into relief the subjective, principled, and pragmatically minded nature of motivated gestural sign formation and interpretation in situated instances of multimodal meaning-making.

Frames, scenes, and metonymy: Seizing the pragmatically minded nature of gesture

When investigating gestures, the focus is naturally on the communicating human body and thus particularly on those frames and experiential domains that involve physical postures and actions that are indexically anchored in the material and

social contexts of communicating and interacting individuals (e.g., Bavelas et al., 1992; Goodwin, 2011; Kendon, 1995; Mittelberg & Waugh, 2014; Müller, 1998; Streeck, 2009; Streeck et al., 2011). According to Fillmore (1982) the term *frame* covers “any system of concepts related in such a way that to understand any of them you have to understand the whole structure in which it fits” (p. 111). This definition pertains first and foremost to lexical semantics, that is, to word meanings and the knowledge structures they entail. Yet Fillmore (1977) emphasizes how frames are conditioned by motivating practices, enactive experiences, interpersonal transactions, contextual affordances, and socio-cultural conventions (see also Ziem, 2014). Since physical actions and gestural communication are part of these situated factors and also shaped by them, it indeed seems worthwhile to apply the notion of frame to gesture (e.g., Dancygier & Sweetser, 2012; Sweetser, 2012). While the FrameNet project has further developed and specified the concept of frames (in terms of sub-frames and semantic roles; see, e.g., Fillmore & Baker, 2009), this paper builds on Fillmore’s notion of frames prior to the FrameNet project. Crucially, it will be argued that gestures recruiting frame structures tend to pick out elements and dimensions of *scenes* in the Fillmorian (1977) sense of the term, that is, aspects of the motivating context of frames.

In light of the semiotic complexity of multimodal meaning construction during face-to-face interaction, the approach advanced in this paper combines several interrelated cognitive, experientialist understandings of frames and metonymy which lend themselves to be applied to both the linguistic and the gestural modalities. These are: Fillmore’s (1982, 1985) notion of frames; Dancygier and Sweetser’s (2014) concept of frame-metonymy; Barcelona’s (2003, 2009) cognitive-functional view of metonymy; Langacker’s (1993) notion of reference points; and Panther and Thornburg’s (2003, 2004) approach to metonymy and pragmatic inferencing. It is maintained that *bodily reference points* may trigger processes of pragmatic inferencing through embodied practices of sensing, creating, and reinstating habitual actions and their material, spatial, and social environments. Building on previous work on the interplay of metonymy and metaphor in gesture (Mittelberg, 2006, 2010, 2013; Mittelberg & Waugh, 2009, 2014), which centrally includes Jakobson’s views on contiguity and metonymy (Jakobson, 1956; Jakobson & Pomorska, 1983; Peirce, 1960), it will be suggested that different metonymic modes may operate on inner and/or outer contiguity relations within frames. With this in view, the paper examines how speakers’ arms and hands, or whole bodies, and the visual action (Kendon, 2004) they perform in multimodal communicative acts may not only provide dynamic iconic structure, but also different kinds of ‘built-in’ indices, both of which may serve as bodily cues for the invocation of basic and more complex frame structures as well as connected metonymic and metaphoric processes of association and signification (Mittelberg & Waugh, 2014).

Given the focus on frames, metonymy, and pragmatic inferencing, the interaction of referential and pragmatic functions assumed by gestures will be of particular interest. To account for the multifunctionality typically exhibited by gestures (e.g., Müller, 1998), a *hierarchical view* (Mittelberg, 2013, p. 755) on gesture functions in multimodal communicative acts is adopted. Combining elements of Peirce's (1960) pragmaticist semiotic theory and Jakobson's (1960) model of speech functions, a hierarchized ordering of semiotic modes and discourse functions in gestural sign processes is assumed. This view entails that, first, the different semiotic modes (e.g., *iconicity*, *indexicality*, and *conventionality*; Peirce, 1960) and speech functions (i.e., *referential*, *emotive*, *conative*, *phatic*, *meta-linguistic*, and *poetic*; Jakobson 1960) are not mutually exclusive. Second, multifunctionality is not solely a matter of mixing or layering functions or dimensions (e.g., McNeill, 2005), but in most gestures we can expect one function to be predominant.¹

Hands framing New York Harbor from Pier 18

As an entry point to frames structures, let us consider the photograph by Shunk and Kender shown in Figure 1, which incorporates many, but not all, of the cognitive semantic and pragmatic mechanisms that are central to this paper.

Hands framing New York harbor from Pier 18 clearly evokes not only the general idea of framing a scene or an event from a certain vantage point, but also an array of (inter-)related frames (Fillmore, 1982, 1985). First of all, we see a literal instantiation of an *embodied frame* formed by two hands occupying the center of the image. The hands are part of the BODY frame, here enacted by the body of the person making the framing gesture. Besides this physically instantiated PICTURE frame superimposed on a real-world scene situated in New York harbor, there are several other frame structures, or various experiential and knowledge domains (Barcelona, 2003; Croft, 1993), that we can identify: for example, the BOAT frame, HARBOR frame, BOAT TRIP frame, BOAT CREW frame, SHIPPING frame, CITY frame, the NEW YORK CITY frame, the EAST COAST frame, the USA frame, as well as the temporal frame constituted by the YEAR OF 1971, entailing how New York and boats looked at the time and what was happening in the US and in the world during that period. Besides these general, shared frames, we can also expect subjective associations metonymically linked to this place, time, and harbor scenario to be invoked in a person looking at the photograph: for example, personal memories of events at the time, experiences with boats, boat trips, harbor

1. For details on this hierarchical account of gesture functions, see also Mittelberg, 2006, 2010; Mittelberg & Joue, 2017; Mittelberg & Waugh, 2009, 2014.



Figure 1. *Hands framing New York harbor from Pier 18*

Harry Shunk & Jonas Kender, 1971, © J. Paul Getty Trust

visits, New York, the U.S., as well as books one has read, movies one has seen, and stories one has been told about these places and these kinds of events, and so forth.

Furthermore, *indexicality* is necessarily a constitutive factor in this iconic representation. Like every photograph, this particular shot of New York harbor has indexical, viewpointed properties (Peirce, 1960): it reflects the angle and perspective from which it was taken. As indicated in the title, the person making the framing gesture and also the photographer were located on Pier 18, from which they had this particular view on the scenario. Gestural portrayals generally tend to reflect character or observer viewpoint (McNeill, 1992, Sweetser, 2012), that is, whether a speaker describes a given scene by personifying another person, e.g., a movie character, or describes the spatial arrangement of a scene from the outside (see Sweetser, 2012, for an overview). Not surprisingly, *metonymy* is also at work in several ways. For instance, there are two nested part-whole (e.g., synecdochic) relationships: first, the photograph represents a part of a larger harbor scene and limits our access to the zone it frames; second, within this image space, two hands create a framed center zone containing the boat and around which the surrounding area of the harbor is visible. Put differently, the superimposed frame configuration causes imposed contiguity relations to arise between the manual frame, the contents it holds, and the elements outside of it (for more details see the section on contiguity relations below).

What the static and silent photograph of two hands framing a New York harbor scene cannot exemplify is that gestures, by their very nature, are dynamic visuo-spatial articulations that usually occur with speech, and that speech and gestures contextualize each other in multimodal discourse. This observation leads us to another building block of the present approach to multimodal discourse pragmatics, namely, the idea of *crossmodal* meaning construction. The speech accompanying a gesture may not only be instrumental in determining which parts and movements of the speaker's body or of particular bodily articulators become metonymically profiled, but also whether the attention stays on the profiled corporeal zones and actions, or *shifts* onto implied imagined entities, persons, or spaces immediately contiguous to, or within the reach of, gesturing hands and signifying bodies (Mittelberg & Waugh, 2014).

In view of the numerous frames that the harbor scenario may evoke, some general questions arise at the juncture of frame semantics and gesture pragmatics, for example, what kinds of frames and experiential domains lend themselves to gestural portrayal? And, if we zero in on frame-internal elements and dimensions, such as associated material artifacts, spatial structures, participant roles, and physical actions and interactions, which of them can actually be portrayed or pointed out by gestures? And which of them typically get enacted through gestures? This paper cannot fully answer these questions. Its aim is to reveal some tendencies in how gestural communicative actions recruit scenes and frame structures.

The paper is structured as follows. The first of four main sections introduces Fillmore's notions of frame and scene and proposes two kinds of frame structures with increasing levels of schematicity and complexity that play a role in gestural frame evocation. This proposal is further refined and theoretically contextualized in the second section. Cognitive linguistic and Jakobsonian perspectives on metonymy and contiguity are discussed in the third section. In the last section, cross-modal mechanisms of pragmatic inferencing in co-speech gesture are discussed in relation to frames and metonymy. The paper concludes with a synopsis and avenues for further research.

Embodied frames and scenes: Theoretical premises and some tendencies in their gestural evocation

Frames, as introduced by Fillmore (1977, 1982, 1985), occupy a central place in various subfields of cognitive linguistics. Frames have been shown to underpin lexical semantics, mental spaces, constructions, metaphor and metonymy, and the understanding of discourse, to name only a few core areas of linguistic structure and meaning construction (see, e.g., Coulson, 2001; Croft & Cruse, 2004;

Dancygier & Sweetser, 2014; Ziem, 2014). In this section, frames are discussed in light of their embodied dimensions, and two interrelated ways in which gestural action may draw on frame structures are proposed (first subsection). Attributing special attention to motivational, conceptual, and contextual factors of frame-based gestural acts, Fillmore's concept of scenes – a central building block of the present view of gesture pragmatics – then is drawn upon to further characterize tendencies in gestural practices of frame evocation (second subsection).

Basic and more complex frame structures: A first characterization

By the word 'frame' I have in mind any system of concepts related in such a way that to understand any of them you have to understand the whole structure in which it fits; when one of the things in such a structure is introduced into a text or into a conversation, all of the others are automatically made available.

(Fillmore, 1982, p. 111)

According to Fillmore, frames are conceptual structures rooted in patterns of experience and practices. Words evoke frames; and frames determine the meaning of linguistic expressions and guide their use in relation to the corresponding motivating experiences.

A 'frame', as the notion plays a role in the description of linguistic meanings, is a system of categories structured in accordance with some motivating context. ... The motivating context is some body of understandings, some pattern of practices, or some history of social institutions, against which we find intelligible the creation of a particular category in the history of the language community.

(Fillmore, 1982, p. 119)

An often-cited example is the WEEK frame (Fillmore, 1982). When a friend emails us saying "Please join us for dinner at the Thai restaurant on the Commons on Friday night at 8 p.m", we understand the temporal information regarding the day against the backdrop of the entire week, that is, the WEEK frame and the cultural background knowledge it implies, for example, that (in most Western cultures) a week consists of seven days, typically grouped into five workdays and a weekend; "8 p.m." receives its meaning in relation to the DAY frame consisting of 24 hours. If we now think of how a gesture could portray the concept 'Friday' in a multimodal description, it becomes clear that, given its abstract nature, characterizing Friday gesturally necessitates seeing it in relation to the larger frame structure of a week, in which it has a clearly defined slot. A gesturer could, for example, lay out a grid in gesture space by pointing to different locations situated next to the other, with each of them representing a particular day of the week

and linguistically highlighting the point representing Friday (see Calbris, 2011, on gestural representations of timelines).

There is an additional frame that the above invitation evokes, namely, the RESTAURANT DINING frame (Schank & Abelson, 1977), which implies elaborated scenarios with certain, culture-dependent sets of elements and sequences of activities (Fillmore, 1982). In the gesture modality, one can evoke individual frames that are part of this larger frame structure by imitating routinized physical actions typically performed in a restaurant setting, such as looking at the menu, trying to get the waiter's attention, eating with a knife and fork or with chopsticks, looking at the plate placed in front of us, searching for one's wallet, signing the check, etc.

So far, the discussion has been based on frame examples that figure prominently in Fillmore's writings and in the literature on frame semantics (e.g., Coulson, 2001; Croft & Cruse, 2004; Dancygier & Sweetser, 2014; Ziem, 2014) and only on hypothetical scenarios and gestural portrayals. These nevertheless allow us, in a first approach, to assign the above examples to two interacting kinds of frame structures that can be assumed to motivate different kinds of gestural practices in multimodal discourse. Gestures based on physical routines, such as eating and paying in a restaurant, are good examples of mimetic communicative actions that may evoke *basic physical action and object frames*, understood as directly experientially grounded frames involving physical action and interaction with the material and social world. Such frame evocations may include intransitive, transitive and ditransitive actions (Hopper & Thompson, 1980).

An example of a gesture evoking a basic physical action frame is given in Figure 2 (adapted from Mittelberg & Waugh, 2014), in which the British architect Norman Foster enacts a fictive scene in an art gallery that he himself designed (*Sperone Westwater Gallery*, Manhattan). Adopting character viewpoint (McNeill, 1992; Sweetser, 2012) Foster mimics how someone inside the building is looking up and pointing to the bottom of an elevator located directly above the gallery's lobby. He is visibly amused by this special architectural feature: "... because the last place you think you'd ever really want to be in any building is underneath the elevator." His words cue this instantiation of an otherwise unspectacular action frame, thus putting a twist on one part of the (more complex) GALLERY VISIT frame, namely, one may find oneself underneath a huge elevator when walking into the gallery. This image of a physical action (i.e., a "body action image icon" according to Mittelberg & Waugh, 2014, p. 1751) portrays, metonymically, only some essential aspects and moments of the full *intransitive* action in question, with several incorporated indices guiding the viewer's attention upward in the direction of the imagined elevator.

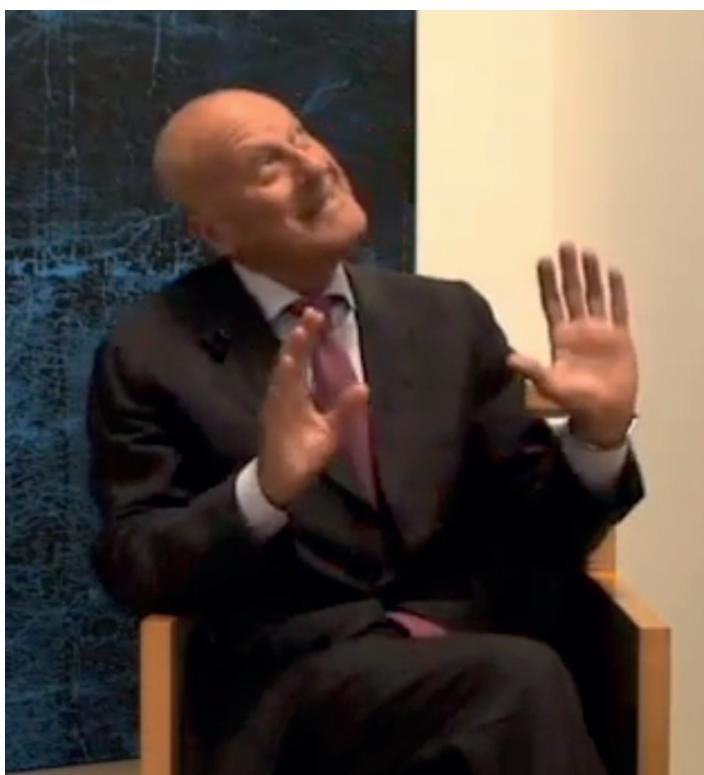


Figure 2. Gestural evocation of a *basic physical action frame* ('underneath the elevator')

Let us now consider a prototypical physical action frame involving both an object and a tool, namely, slicing an apple (Figure 3). In a study on *transitive* action gestures (Grandhi, Joue, & Mittelberg, 2011), speakers describing everyday actions showed a tendency to produce gestures in which the dominant hand pretends to be handling the tool required to perform a particular action. While explaining "you need to slice the apple by holding it down and cutting it there," one participant acts as if she were using a knife (tool) with her right hand, as shown in Figure 4, while imitating a cutting action on an imaginary apple (object) that she is seemingly holding down with her left hand. While the speech content draws attention to the action and the object, the tool remains unmentioned. However, in the context of this particular frame with very strong, embodied connections between its prominent parts, the knife is pragmatically, that is, *automatically* inferred (see Fillmore, 1982, p. 11, quoted above; Dancygier & Sweetser, 2014).

While involving a sequence of basic actions and objects, Fillmore's (1982) example of the RESTAURANT DINING frame and the WEEK frame illustrates what is meant by more *complex, highly abstract frame structures*, that is, those pertaining

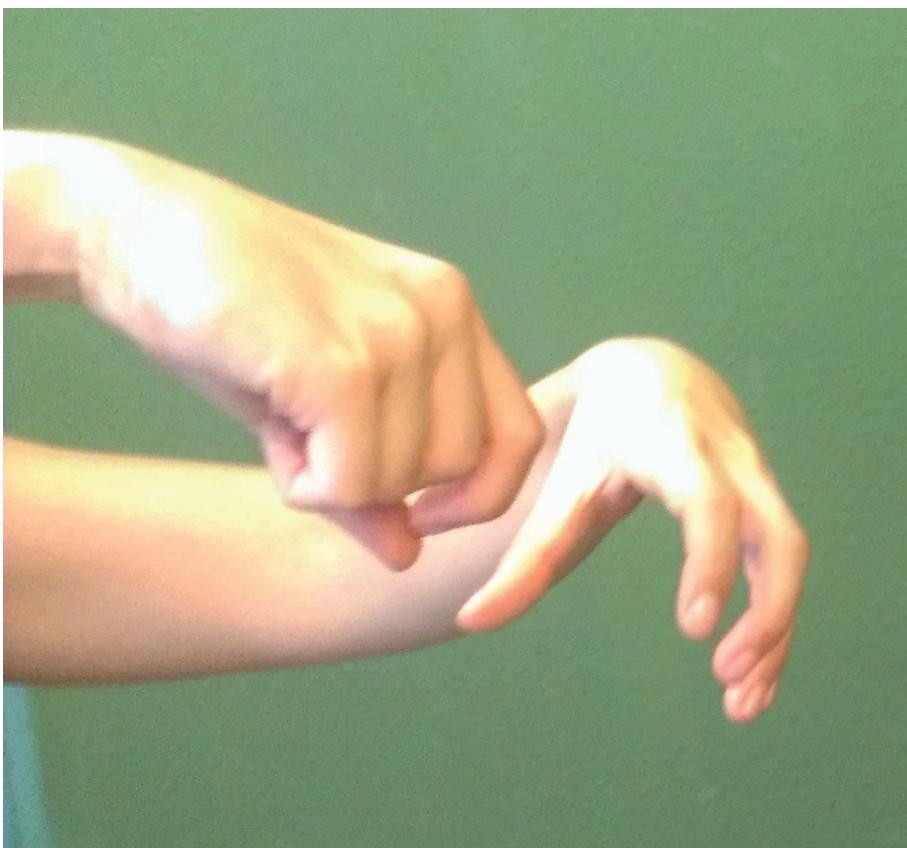


Figure 3. Gestural evocation of a *basic physical action and object frame* ('slicing an apple')

to abstract knowledge structures. As just mentioned, individual physical action units may be embedded in larger, more complex frames or scripts. A frame displaying an even higher level of abstraction is the SENTENCE frame instantiated by the line-drawing gesture shown in Figure 4a (adapted from Mittelberg, 2006). The speaker, a linguistics professor, says the following: "So we think of a sentence as a string of words ... it's obvious that we can group the words together in various ways." After setting up a minimal reference structure in gesture space (Figure 4a), namely, an imaginary piece of line representing a string of words, the speaker uses both of her hands to seemingly group elements in a sketchy way along the imaginary string (Figure 4b), moving them along the string from left to right.

This kind of dynamic diagrammatic representation of chunked word groups within a SENTENCE frame was probably motivated by the fact that the speaker had just explained *bracketing*, a classification system by which elements in a sentence are grouped into constituents, each of which is enclosed in square



Figure 4a. Gestural evocation of *complex, highly abstract frame structures* ('a string of words')

brackets (with less complex linguistic units being embedded in more complex units). Gestural graphs and diagrams generally are highly abstract schematic representations that may bring out the internal structure of a frame by highlighting individual components or slots, the boundaries between its parts or how the elements are connected within a larger whole (see Mittelberg, 2008, for gesturally evoked hierarchical tree structures). In the following, Fillmore's concept of scene will be operationalized to further specify these first characterizations and to then draw (in the following main section) a more detailed picture of frame-based gestural action and communication.

Scenes and their relevance for gesture

One of the central assumptions put forward in this paper is that coverbal gestures that recruit frame structures have a particular and natural propensity to pick out elements and dimensions of scenes in the Fillmorian sense of the term. In the context of gestural frame evocation a crucial point is that while frames are abstract conceptual structures, they are clearly "motivated by human experiences, social institutions, and cultural practices" (Coulson, 2001, p. 18). Fillmore (1977, 1985) counts extra-linguistic, contextual, and co-textual factors among the forces that shape the understanding of an expression as well as the category systems that constitute frames. Let us take a brief look at some of his earlier writings where, in addition to frames, he uses the term *scenes* to encompass real-world experiences,



Figure 4b. Gestural evocation of *basic physical action frame* ('group words together')

situative factors, and non-linguistic conceptual dimensions (e.g., Fillmore, 1975; for a detailed discussion see Ziem, 2014, pp. 24, 188ff.).

Generally speaking, scenes are particularly relevant with respect to how interlocutors construe and follow processes of online meaning construction: "In most natural conversations, the participants have, already 'activated', a number of shared, presupposed, scenes that we can speak of as being in their consciousness as they speak" (Fillmore, 1977, p. 126). It thus seems safe to assume that scenes factor into the dynamic contextualized processes that shape both the production and interpretation of gestures with speech. Regarding the gradual abstraction processes during the acquisition of word meanings, Fillmore describes the role of situative settings in the emergence of different kinds of scenes as follows:

It appears ... that in meaning acquisition, first one has labels for whole scenes, then one has labels for parts of particular familiar scenes, and finally one has both a repertory for schematic or abstract scenes and a repertory of labels for entities or actions perceived independently of the scenes in which they were first encountered. (Fillmore, 1975, p. 127)

From the following passage, in which Fillmore spells out the difference between scenes and frames, we can conclude that scenes feed into frames, and that frames here are mainly used to capture linguistic choices pertaining to lexical semantics, grammatical rules or grammatical categories:

(P)people, in learning a language, come to associate certain scenes with certain linguistic frames. I intend to use the word *scene* ... in a maximally general sense, to include not only visual scenes but familiar kinds of interpersonal transactions, standard scenarios, familiar layouts, institutional structures, enactive experiences, body image; and, in general, any kind of coherent segment, large or small, of human beliefs, actions, experiences, or imaginings. I intend to use the word *frame* for referring to any system of linguistic choices (the easiest being collections of words, but also including choices of grammatical rules or grammatical categories – that can get associated with prototypical instances of scenes).

(Fillmore, 1977a, p. 63)

The present view on gestural frame evocation thus rests upon the claim that gestures may profile essential aspects of the experiential grounding of scenes, that is, the motivating contexts of frames. For instance, they may metonymically abstract the onsets of physical actions or salient parts of the social and material culture that constitute the foundation not only of basic frames, but also of more schematic and highly abstract dimensions of our experience and embodied knowledge of the world. In language, grammatical constructions exhibiting different degrees of complexity belong to the abstract, schematic structures that draw on and reflect “scenes basic to human experience” (Goldberg, 1995, p. 5; see also the section on complex frame structures below).²

Towards a frame-based account of gesture pragmatics

Based on the foregoing preliminary discussion of frames and scenes, we are now in a position to further refine our understanding of how they may be recruited through gesture in certain utterance contexts. This will be done by drawing connections to related theoretical notions. It needs to be kept in mind that the account developed here characterizes certain tendencies in gestural practices of evoking interlaced, dynamic experiential structures of different levels of embodiment, schematicity, and complexity. These patterns are not to be seen as categorical or exhaustive.

Basic physical action and object frames

In the present account, *basic physical action and object frames* roughly correspond to the first repertory defined by Fillmore (1975) as encompassing “schematic or

2. Fillmore (1985) ceases to use the term *scene*; however, the motivating, situated aspects remain part of his understanding of frames.

abstract scenes” (p. 127; see full quote in the section on scenes above). They pertain to actions and entities that are perceived in connection with the scenes in which they were encountered and embodied during language acquisition. For these early stages of development, Zlatev’s (2005) notion of mimetic schemas, such as GRASP, JUMP or CRAWLING or FLYING, as well as Andrén’s (2010) notion of *action gestalts* seem pertinent.³ Basic physical action frames further seem akin to what Lakoff (1987) refers to as *basic-level actions*: “we have basic-level concepts not only for objects but for actions and properties as well. Actions like *running*, *walking*, *eating*, *drinking*, etc. are basic-level” (pp. 270–271, italics in the original). Such basic-level actions can be easily enacted through manual gestures, head movements, or the whole-body (e.g., Figures 2 & 3). For instance, Mittelberg & Rekittke, in press, describe a gestural swimming action performed to illustrate the SWIM-WEATHER frame.

Other constructs of relevance here are *primary scenes* and *subscenes* (Grady & Johnson, 2002), which are a means to break down experience types into their component dimensions. For example, regarding the primary scene, or recurring experiential scenario, “manipulation of complex object” (p. 537), gestural practices of object manipulation can draw on specific subscenes, such as hands carefully holding a brittle object, exploring its surface, taking it apart, etc. By the same token, basic action and object frames seem to comprise what Slobin (1985), in his work on the acquisition of grammatical categories, calls *prototypical events*. These also include object manipulation and object transfer, as well as “grasping, pushing, squeezing, pulling apart, etc., each of which would be a distinct subscene” (Grady & Johnson, 2002, p. 548). In the study on transitive action types mentioned earlier (Grandhi et al., 2011), participants demonstrated various ways of slicing an apple, that is, they were drawing on distinct subscenes of the primary scene of dividing an object into pieces. An alternative practice to the one shown in Figure 3 was performed with a flat vertical right hand signifying the blade of a knife that is cutting an imaginary apple sitting on the open left hand which needs to be pragmatically inferred.⁴ Such schematic or abstract scenes may serve as blueprints for gestures simulating action that involve the whole body or parts of the body, as posited by Hostetter and Alibali’s (2008) gestures-as-simulated action framework, which

3. Zlatev (2005, p. 334) defines mimetic schemas as “dynamic, concrete and preverbal representations, involving the body image, accessible to consciousness and pre-reflectively shared in a community”; they are situated at an intermediate level of abstraction (Zlatev, 2013; see also Cienki, 2013b). See Mittelberg & Joue, 2017, for a more detailed frame-based discussion of image schemas, mimetic schemas, and metaphor in gesture.

4. See Grandhi, Joue, & Mittelberg (2011) on applying these basic transitive action routines to the realm of gesture-based interactive systems.

asserts that “gestures emerge from the perceptual and motor simulations that underlie embodied language and mental imagery” (p. 502; see also Gibbs, 2006).

When basic physical actions and the implied objects receive a metaphorical meaning through the local speech context the tight correlation between certain physical experiences and conceptualization becomes very apparent (e.g., Cienki, 1998, 2013; Cienki & Müller, 2008; Müller, 1998; Sweetser, 1998). Grady (1997) coined the term *primary metaphor* for metaphorical mappings rooted in such tight experiential correlations (e.g., CATEGORIES ARE CONTAINERS or AFFECTION IS WARMTH). So, naturally, basic action and object frames may also be evoked through gestures when speakers talk about abstract ideas and processes. Through enacting embodied image-schematic, force-dynamic, and metaphorical dimensions of the properties or behavior of, for instance, verbs, atoms, or numbers, speakers may become, hold, show, or otherwise manipulate imagined abstract notions and thus learn, or teach, something about them (e.g., Calbris, 2003; Mittelberg, 2008; Núñez, 2008; see also Mittelberg & Joue, 2017, for a frame-based account of metaphorical construals involving gestural action).

For an example of a metaphorically construed gestural portrayal, consider the following basic action of seemingly holding an object with two hands (Figures 5a–c). During a first-year German language course, the instructor shown here invites the students in the target language to reflect on the properties of the grammatical category *infinitive*: “You are looking at an (...) infinitive, can you say.. ah, this is (...) present, perfect, or past perfect (...)?” (German original: “Sie schauen einen (...) Infinitiv an, können sie sagen.. ah, das ist (...) Präsens, Perfekt, Plusquamperfekt (...)?”).

In the moment captured in Figure 5a, the speaker seems to be holding some kind of imaginary generic object with her two hands and is mimicking what she is asking her students to do, that is, to look at the grammatical category ‘infinitive’ that her words are referring to and that is metaphorically represented here as a tangible object. She then directs her gaze upward and toward her audience (Figure 5b), still holding the bimanual gesture while asking the students whether infinitives may be marked for tense. In Figure 5c, while she is waiting for a response from the students, her right hand forms into a palm-up open hand, PUOH (Müller, 2004), and her left hand also opens up slightly. She stays like this for four seconds without saying anything. So while in her speech she tries to help the students develop a solid understanding, or a frame, of infinitives, her gesture is drawing on the basic scene of holding an object with two hands.

This multimodal portrayal unites referential and pragmatic functions, with the latter becoming gradually more foregrounded: first the speaker enacts the speech content while also adopting the students’ viewpoint (Figure 5a), then she directly addresses the students by asking a question (Figure 5b), and finally, the gesture in



Figure 5a. Looking at *infinitive*



Figure 5b. Asking a question about *infinitive*



Figure 5c. Waiting for an obvious answer

Figure 5c is predominantly pragmatic and interactive (i.e., emotive, conative and phatic, according to Jakobson, 1960; see Mittelberg & Joue, 2017, for more details). The PUOH can be taken to indicate that the metaphorically construed, reified infinitive is still available for inspection (just in a reduced version) and/or that the speaker is waiting for the answer which she herself provides a moment later (i.e., no, infinitives are not marked for tense). The gestures' semantic and pragmatic dimensions are reinforced by the speaker's facial expressions, which change from attentively inspecting the object to interrogative (with slightly raised eyebrows), and finally, her frowning eyebrows signal that it quite obviously seems very unlikely that an infinitive exhibits tense marking.

Higher levels of action-based abstraction may generally result in gestures' predominant pragmatic functions, as found in certain *gesture families* (Calbris, 2011; Kendon, 2004) or *recurrent gestures* (e.g., Ladewig, 2014), such as the family of *away gestures* described by Bressem and Müller (2014). Wehling (2010) distinguishes between *agonistic offense* and *antagonistic defense gestures* based on primary metaphors (see also Müller, this issue, and Wehling, 2017). Moreover, basic manual actions – such as holding something or giving something to somebody – have been attested a certain grammatical affinity regarding prototypical instances of linguistic transitive and ditransitive constructions (e.g., Goldberg,

1995). It makes sense, then, that such routinized manual actions may function as the substrate of embodied grammaticalization processes in gesture in which the correlated frames and scenes as well as (primary) metaphors and metonymy interact. For instance, basic schematic scenes recruiting actions of giving or holding may feed into non-complex constructions, as Mittelberg (2017) has argued concerning intransitive multimodal existential constructions in German ('es gibt', Engl. 'there is') involving intransitive linguistic uses of the originally ditransitive verb 'geben' ('to give') and palm up open hand gestures (see also Streeck, 2009, and the following subsection).

Complex, highly abstract frame structures

Complex, highly abstract frame structures encompass schematic, larger frame networks assumed to partake in human backstage cognition, thus structuring, for instance, the conceptual system, abstract knowledge domains, and entire discourses (e.g., Coulson, 2001). They are akin to the "repertory of labels for entities or actions perceived independently of the scenes in which they were first encountered" (Fillmore, 1975, p. 127; see full quote in the section on scenes above), which are supposed to mostly relate to linguistic structures and grammatical categories. This section is concerned with such composite constructs, for instance, linguistic and discourse structures as well as complex (non-linear) knowledge systems.

As briefly addressed in the previous section, in gesture research the relation between grammar and gesture has been approached from different angles. For instance, work on meta-linguistic and meta-grammatical gestures produced by linguists while teaching has provided insights into their mental models of language and grammar, that is, how they metaphorically conceptualize single linguistic categories, drawing on what in this paper are called basic action and object frames, but also larger, internally structured units such as phrases and sentences, as well as linguistic theories (Mittelberg, 2006, 2008). A growing body of research into the multimodal nature of grammar further suggests that gestures may not only assume functions comparable to those of verbs, nouns, and adjectives (e.g., Fricke, 2012), but that they also partake in certain multimodal constructions (e.g., Andrén, 2014; Schoonjans, 2014; Schoonjans et al., 2016; Benazzo & Morgenstern, 2014; Steen & Turner, 2013; Zima & Bergs, 2017).

Since the organization of abstract knowledge domains generally relies on very high levels of schematicity, the kinds of complex frame structures of interest here are more removed from the respective motivating contexts of experience (Fillmore, 1975). They function as ordering principles of embodied conceptual structures and processes such as image schemas, force dynamics, blends, and general metaphoric mappings like ABSTRACT STRUCTURES ARE PHYSICAL

STRUCTURES (Sweetser, 1998; Dancygier & Sweetser, 2014; Parrill & Sweetser, 2004). In the gestural modality, such larger schematic architectures may or may not be overtly represented. Speakers can make them visible in the form of virtual line drawings (e.g., the ‘string of words’ discussed above, Figure 4b), configurations of locations pointed to in gesture space, or other kinds of incrementally emerging gesture complexes that highlight how individual actions, items, places, people, events, concepts, or discourse contents relate to one another (e.g., Bressem, 2012; Enfield, 2011). In addition, through meta-discursive gestures, such as beats (McNeill, 1992), speakers may accentuate individual elements in their utterances, thus making them stand out from the larger composite structure. Discourse-structuring gestures further serve to group discourse contents at a meta-level, for instance, by systematically placing them in certain segments of gesture space (see also Müller, 1998; Streeck, 2009).

Let us take a closer look at complex, highly abstract frame structures that are assumed to operate in the background of cognition and communication and that may provide some sort of dynamically evolving cognitive-semiotic structure along which speakers may get from one point to the next in their imagination, reasoning, and discourses (Coulson, 2001; Dancygier & Sweetser, 2014; Fillmore, 1977). Storylines, films, plays, rituals, common scenarios (e.g. scripts), and discourse genres belong to these larger frames with numerous parts, participant roles, rich inferences and imagistic textures (e.g., Labov, 1972). Based on habitual, structured semiotic experience provided by, for example, genres or routines, they enable viewers or interlocutors to identify deviances and anticipate what comes next. The *Canary Row* episodes (e.g., McNeill, 1992, 2005), for instance, are organized by scripts and frames people can fall back on when recounting these animated cartoons. The individual character frames, such as those of Tweedy Bird and Sylvester, as well as single scenes, are constitutive parts of those larger frames that can be assumed to also structure the multimodal retellings and thus determine the order in which single basic action frames, like those involved in particular motion events, are gesturally enacted. This entails that larger, overarching frame structures can be evoked in at least two ways: by simulating actions one by one in a sequence or by gesturally representing the sequential structure per se.

Consider the following example of a complex frame structure taken from a study examining how mental frames and mental maps interact when interlocutors jointly plan a trip through Europe and thereby draw travel itineraries in the air (Mittelberg & Rekittke, in press). Conversational partners were found to first suggest alternative routes by gesturally pointing out and connecting various destinations, thus evoking within the EUROPE frame individual city and country frames. For instance, the interlocutors shown in motion-capture stick figures in Figure 6 decided to first travel from Stockholm to Prague, then to Southern Europe and

finally back to the North. The gestural diagram drawn in the air by the speaker on the right visually resumes the travel itinerary finally agreed upon. Note that the concurrent description (in German) can only be understood when considering the preceding discussion and gestures: "ich denke wir fahren dann so weißte [von (-) da nach da rüber runter und dann nach da und dann wieder nach Hause]" (in English: "I think, then we travel like this you know [from (-) there to there across down and then to there and then back home again]"). Against the backdrop of the EUROPE and INTERRAIL frames, this diagram represents the dialogically negotiated, personalized train trip.

As the foregoing discussion demonstrates, gestural portrayals reflecting aspects of larger abstract frame structures – like those underpinning, e.g., discourses, films, theories, maps, time lines, or sentence models – may exhibit high-order iconicity such as diagrams (Peirce, 1960, p. 135; Bressem, 2012; Fricke, 2012; Mittelberg, 2006, 2008, 2010a, 2014) and/or image-schematic structures (e.g., PATH; LEFT/RIGHT; Johnson, 1987). Additional frame-based research is needed to better understand in what ways gestures not only underpin or partake in multimodally achieved utterances, or constructions, but also drive new ideas, story-lines, paths of reasoning, and so forth.

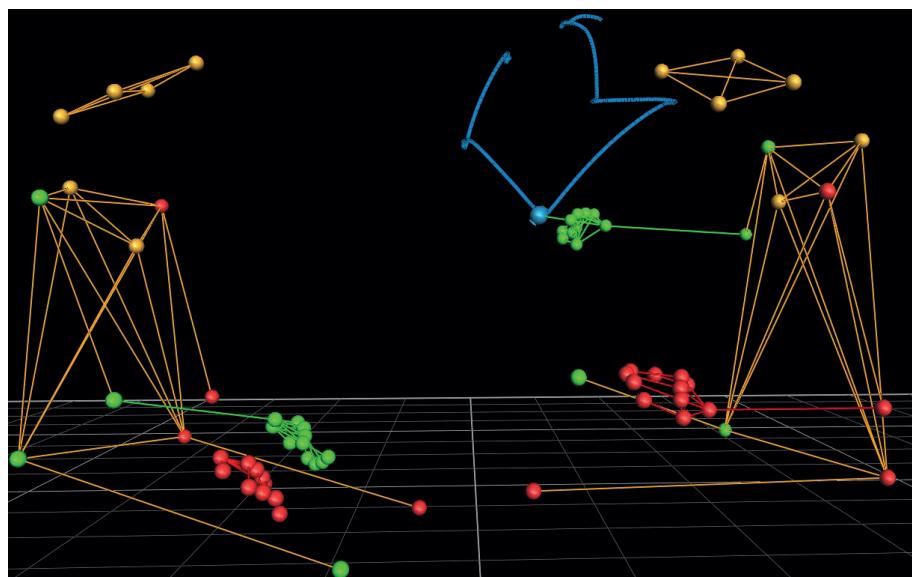


Figure 6. Gestural diagram of travel itinerary evoking INTERRAIL and EUROPE frames

Summary of observations and proposals

Attesting scenes (Fillmore, 1977) a constitutive role in gestural frame evocation, and assuming “the development of a complex frame out of correlated simpler frames” (Dancygier & Sweetser, 2014, p. 138), the following two interrelated types of frames are assumed to motivate and structure, at least to some degree, embodied cognitive-pragmatic processes of multimodal meaning-making, notably the formation and interpretation of frame-evoking, co-speech gestures:

A. *Basic physical action and object frames (AOF)*

- Are experientially grounded, simple frames involving the human body and its routinized physical actions, interactions with the material and social environment, as well as metonymically correlated items and actions (strong experiential correlation);
- Include simulated actions, metonymically abstracting aspects of, e.g., scenes, primary scenes and subscenes, and basic-level actions;
- Exhibit varying degrees of abstraction and schematicity;
- Are typically embedded in more complex frame structures (see (B) below);
- Examples include simple habitual *intransitive* actions, such as walking or looking at someone/something (Figure 2); *transitive* actions, e.g., object manipulation or tool use (Figure 3); interactions between interlocutors, e.g., object transfer; and more abstract patterns, e.g., certain recurrent gestures and gesture families;
- May incorporate image schemas and force-dynamic gestalts;
- May build the basis for metaphorical projection, e.g., primary metaphors;
- May build the substrate for embodied lexicalization, grammaticalization, and basic schematic scenes factoring into constructions (linguistic/gestural/multimodal);

B. *Complex, highly abstract frame structures (CFS)*

- More schematic, larger abstract knowledge structures mainly operating in the background of cognition and communication (weaker/indirect experiential motivation), based on, e.g., visual perception, encyclopedic knowledge, semiotic structures and practices (e.g., drawing, writing);
- Metonymically organized, composite part-whole structures with internal relations providing entry points/slots/boundaries that can be highlighted and operationalized through gestures;
- Typically consist of several correlated simpler elements and/or *basic action and object frames* (see (A) above); e.g., the grouping gestures along the imagined sentence string in Figure 4b);

- Examples include scripts (e.g., the RESTAURANT DINING frame) and abstract knowledge structures such as time systems and theories (e.g., the WEEK frame, SENTENCE frame);
- Manifestations may exhibit diagrammatic iconicity (e.g., graphs, maps, and internally structured image schemas, e.g., SOURCE-PATH-GOAL);
- Include discourse structures, narrative patterns, conversational patterns, film sequences, etc.;
- May manifest themselves in gesture or only provide an underlying cognitive-semiotic grid for multimodal discourse;
- May build the basis for metaphoric structural mappings and semantic networks;
- May build the substrate for embodied grammaticalization and internally structured constructions (linguistic/gestural/multimodal).

In light of the above observations it further seems necessary to distinguish not only between different kinds of frame structures, but also between different kinds of frame-evoking gestural practices. While talking, people not only simulate *basic physical actions* such as running, eating or holding an object), but also *semiotic practices* resulting in some kind of notation: e.g., drawing or writing on paper, on blackboards, or in the sand. By gesturally imitating such semiotic practices, that is, acquired cultural techniques, speakers may compensate for the lack of direct experiential correlation. Drawing gestural paths in the air or seemingly placing several items next to each other in gesture space here can be understood as pragmatically motivated *semiotic grounding practices* that allow speakers to map out a meaningful assembly or figuration, thus revealing parts of the underlying more complex and abstract frame structures. Gestural practices of tracing, placing or grouping items – including sorting discourse contents at a meta-level – are typically performed without being referred to in speech (see Figure 4b). Crucially, such meta-actions seem to reflect the speakers' on-line thinking and construal at a higher cognitive level involving abstract structures, relations, and processes (e.g., Bressem, 2012; Calbris, 2011; Goodwin, 2007; Kendon, 2004; Mittelberg, 2006). It would be interesting to further investigate whether the different gestural practices of frame evocation proposed here pattern with particular viewpoint strategies, for instance, enacting basic action frames with character viewpoint and drawing lines and diagrams of more complex frame structures with observer or dual viewpoint (e.g., Parrill, 2009; Sweetser, 2012).

It is important to keep in mind that neither the frame types nor the gestural practices discussed in this paper can give an exhaustive picture of these complex mechanisms and of the various ways in which they may be instantiated across situations, individuals, languages, contexts, and cultures. Rather, this proposal is

to be seen as a first attempt to map out certain tendencies in gestural embodiment of frame elements and of correlated experience and associations, which may be metonymic and/or metaphoric in nature.

Body-based metonymy and gesture pragmatics

Metonymic relations have been touched upon at various points throughout this paper. In this section, pragmatic functions of metonymy in multimodal discourse take center stage. Given the thematic focus of this special issue, the main aim here is to provide multimodal evidence for the claim that “metonymy is a central organizing principle of pragmatics, the contextual use and interpretation of meaning” (Dancygier & Sweetser, 2014, p. 162). Building on previous work on metonymy in gesture (Mittelberg, 2006, 2010b; Mittelberg & Waugh, 2009, 2014), the aim pursued here is to show how an embodied approach to metonymy may contribute to a frame-based understanding of multimodal discourse pragmatics. Making the human body the starting point for studying the embodied mind (e.g., Gibbs, 1994, 1999, 2006; Johnson, 1987, 2007; Lakoff, 1987), its structure and actions are of particular interest. Here we are concerned with the contiguity relations between the speaker’s body and the physical and imagined world at her/his fingertips that can be operationalized for gestural meaning-making. We will now briefly review those approaches to metonymy that are particularly relevant to the present discussion, each time pointing out how the corresponding concepts may be instantiated in gesture and how these principles may help elucidate gesture pragmatics.⁵

Cognitive/experiential/functional DOMAINS

Domain-based definitions of metonymy hold that whereas metaphor involves a cross-domain mapping – as in *she grasped the new concepts without difficulty*, which involves a projection from physically seizing an object to the abstract domain of understanding – metonymic mappings occur within the same cognitive or experiential domain or idealized cognitive model (ICM; Lakoff, 1987; Panther & Radden, 1999, pp. 19ff.). According to Barcelona’s (2003) functional definition, “(a) metonymy is a mapping of a cognitive domain, the source, onto another domain, the target. Target and source are in the same functional domain and are

5. Cognitive linguists have developed various accounts of metonymy, which for reasons of space cannot be covered in great detail here. See, e.g., Benczes et al., 2011; Dirven & Pörings, 2002; Feyaerts, 1999; Gibbs, 2004; Littlemore, 2015; Panther & Radden, 1999; Panther & Thornburg, 2003; Panther et al., 2009.

linked by a pragmatic function, so that the target is mentally activated” (p. 83). For instance, in *Would you like another cup?* employed to ask a person whether she would like more tea, the word *cup* stands for its content, with container and content being pragmatically tightly linked within the same experiential or functional domain of drinking tea out of cups.

In gesture, body-centered and action-based source domains naturally activate pragmatic links to experientially connected metonymic targets, for example, gesturing hands and motor patterns may evoke the tools or objects they routinely handle when they actually perform actions. For instance, to order another cup of tea in a café, one may raise one’s empty tea cup and signal to the waiter that one would like to have the same hot beverage again by pointing at the cup with the other hand. Drawing on Langacker (1987), Croft (1993) extended the single-domain approach to the notion of *domain matrix*, in which one domain can stand for another through a shift in foregrounding. We can thus say that in the tea cup example, in the café scene, first the cup is foregrounded due to its perceptual prominence (*vis-à-vis* the hot beverage which either has already been consumed or is being ordered, i.e., it is not present in the cup); then, the metonymic process entails a shift in foregrounding onto the metonymic target, that is, the tea indirectly referred to.

FRAMES and frame metonymy

Frames (Fillmore, 1982, 1985) are metonymically structured wholes in which one of its parts may evoke another part or the frame as a whole. So there are frame-inherent networks of metonymic relations and associations which may be activated and operationalized in processes of frame-based reasoning, language use, and discourse understanding (e.g., Dancygier & Sweetser, 2005, 2014). In the introduction to this paper, an array of frames evoked by the photograph *Hands framing New York harbor from Pier 18* (Figure 1) was identified: the BOATING frame, the HARBOR frame, the NEW YORK CITY frame, etc. Each frame offers various springboards for metonymic associations as well as entry points and conceptual bridges for (collaborative) story telling. In ongoing conversations, speakers may not only frame a certain scene from a certain perspective and in positive or critical terms, but also use linguistic, gestural, and eye-gaze cues to evoke semantic frames. Hence, alluding verbally or gesturally to a particular discourse-relevant element may automatically trigger connections to directly correlated elements, the frame as a whole, or more far-reaching metaphoric associations. To return to the email scenario, if I imitate typing with both hands while saying to a friend “I’ll email you after the meeting”, the mimicked typing action takes place on an imaginary keyboard. Due to the verbal cue, the pantomimed action of typing not

only gets profiled against the ground of the imaginary keyboard, thus evoking the TYPING frame, but also against the backdrop of the EMAIL frame as a whole.

The email example is a multimodally achieved instance of what Dancygier and Sweetser (2014, p. 134) call *frame metonymy* (as opposed to *categorial metonymy*; *ibid.*). The authors define metonymy as “the use of some entity A to stand for another entity B with which A is *correlated*” (*ibid.*, italics in the original). They further distinguish between a more general concept of *frame metonymy* referring “to all usages where one reference to an element of a frame is used to refer to either the frame as a whole or to other associated elements of the frame” (p. 135; e.g., ‘the Crown’ referring to the British monarchy), and *part-whole frame metonymy* (e.g., ‘field hands’ standing for persons who mainly do physical work). With reference to Langacker’s (1987) notion of *active zone*⁶ as the profiled part of a whole, Dancygier and Sweetser (2014, p. 108) stress the fact that a certain degree of salience is needed to clearly associate a term with a frame. As for part-for-whole frame metonymy, “the part centrally or directly involved in an activity stands for the whole. The hand, for example, is the part of the arm used for holding, touching, etc.; hence it is the active zone of the arm for many purposes” (p. 144). This certainly applies to gesturing hands as the active zone of the gesturer’s body that stand for the entire person. Hands and arms are the parts of the body that are directly involved in both physical actions and communicative gestures; however, they do not necessarily evoke the entire body when gesturing, but may also engender semiotic processes in their own right (Mittelberg, 2013; see also Hassemer, 2016). In such cases, the following metonymic mappings may be said to apply: SALIENT PART OF FORM FOR WHOLE FORM (Barcelona, 2009, p. 395) and SALIENT PHASE OF ACTION FOR WHOLE ACTION. Object-oriented gestures (as in Figures 3, 4, & 5) may further reflect the following metonymic correlations: ACTION-FOR-OBJECT INVOLVED IN ACTION, PRESENTATION-FOR-PRESENTED, OR LOCATION-FOR-OBJECT.

Much of what has been described above in terms of cognitive, experiential, or functional domains can thus also be understood in frame-semantic terms (see also Panther & Radden, 1999, p. 9). As Dancygier and Sweetser (2014, p. 102ff.) convincingly show, compared to domains, the structural organization of frames allows for a more systematic and fine-grained account of correlations both within a frame (thus giving rise to frame metonymy) and between two frames that are partially mapped onto each other (thus giving rise to metaphor). Central to the present perspective on frame-based gesture pragmatics is their point that:

6. Langacker (2009) defines *active zones* as follows: “An entity’s active zone, with respect to a profiled relationship, is that facet or fit which most directly and crucially participates in that relationship” (p. 48).

(f)rame metonymy is closely tied to the kind of correlations which are involved in experientially based metaphors, in particular Primary Metaphors (...). It is precisely the development of a complex frame out of a correlated simpler frame which makes a primary scene so powerful. (Dancygier & Sweetser, 2014, p. 137)

Indeed, evoking the more complex EMAIL frame through a simple gestural action (i.e., typing) allows the addressee's embodied mind to associate not only experientially grounded connections, such as a physical keyboard or touchpad typically used to type on, but also more immaterial elements, such as subsequent phases of the process, possible contents of the email message, her own reply, as well her friend's anxious anticipation of it. So it seems that frame-metonymic associations are likely to also have subjective dimensions that may activate emotional or mental states, such as joy, anxiety, or uncertainty, that is, what Johnson (2005) calls "felt qualities of our experience, understanding, and thought" (p. 31; see also Mittelberg, 2013a/b).

As for the tea example, put in frame-semantic terms, the word *cup* operationalizes an entrenched correlation between the physical elements, actions, and cultural conventions that together constitute the TEA-DRINKING frame. We can also say that there is a *metonymic shift* occurring within the frame structure from the container 'cup' to the desired content 'tea' (Panther & Radden, 1999, pp. 9ff., see also Coulson, 2001; Fauconnier & Turner, 2002). Cultural conventions as to how tea is prepared and served can be expected to influence the ways in which speakers gesturally imitate single actions, drawing on primary scenes, and sub-scenes (Grady & Johnson, 2002; see also the section on scenes above) that constitute and are interrelated within the more complex frame based on particular customs. Consider, for instance, the differences in preparing and serving tea in Western cultures as compared to Japanese tea ceremonies.

Contiguity relations operationalized in gesture

Generally speaking, there are innumerable contiguity relations out there in the world, in our imagination, and in our embodied knowledge structures that may (or may not) be operationalized for reasoning and communication. For instance, in Figure 1 there are contiguity relationships between the pier and the boat, Pier 18 and the other piers, the boat and the Atlantic ocean, the boat and the people on it, etc. The two hands forming the superimposed frame are also contiguous through their touching fingertips. The question that remains is how do metonymic processes within frames draw on such latently meaningful junctures where metonymic shifts may take place; that is, how can one entity be used to refer to a contiguous, correlated item (Dancygier & Sweetser, 2014)? Moreover, how do speech and gesture interact in constructing meaning at such transition points?

In the cognitive linguistic literature, contiguity relations feeding into metonymic expressions are seen as either objectively given or cognitively construed (e.g., Dirven & Pörings, 2002; Panther & Radden, 1999). They are further assumed to be contingent (Panther & Thornburg, 2003), which seems pertinent for bodily semiotics since a gesture is, in most cases, just a gesture because gesticulating hands do not manipulate physical objects or surfaces; they only pretend to do so. For instance, in the email-typing gesture the original contiguity relation within the frame of hands typing on a keyboard is cancelled (Mittelberg, 2013). Peirsman and Geeraerts (2006) put forward a prototype approach to conceptual contiguity and metonymy. Taking the spatial and material domain as the prototypical core of contiguity, they posit a continuum of strength of contact as the basis for spatial, temporal, as well as abstract domains (including events, actions, processes, and assemblies). In the spatial/material domain, the continuum extends from *spatial part and whole* (e.g., head/person) to less prototypical cases, such as *containment/container* (e.g., milk/glass), *location/located* (e.g., house/inhabitants), and *entity/adjacent entity* (e.g., person/clothing).

Let us now turn to earlier work by Roman Jakobson, whose account of contiguity relations has proven insightful when it comes to describing the experiential and semiotic motivation for metonymy in multimodal discourse. In his writings on aphasic disorders Jakobson (1956) not only showed how deeply rooted the distinction between similarity (metaphor) and contiguity (metonymy) is, but he also went on to differentiate contiguity relations in the physical world, for example, between a knife and fork, and those which combine items in a semiotic contexture, for example, linguistic units forming syntagms or entire discourses (Dirven & Pörings, 2002; Waugh & Monville-Burston, 1990). Moreover, Jakobson distinguished between two fundamental contiguity relations, namely, inner and outer contiguity, which in his understanding feed into distinct metonymic operations (Jakobson & Pomorska, 1983; as applied to gesture, Mittelberg, 2006, 2010a, 2013; Mittelberg & Waugh, 2009, 2014).

Inner contiguity, Jakobson asserted, underpins part-whole relationships, or synecdoche, that is, a part stands for another part, a part for the whole, or a whole for the part. *Internal metonymy* operationalizes these kinds of contiguity relations inherent to a given gestalt. For instance, in the expression “everyone lives under one roof”, ‘roof’ evokes the entire house of which it constitutes a physical fragment. Internal metonymy entails that the inner structure of a body, entity, or action is broken down into its component parts or phases, and that one of them is taken to imply a connected component or the entire gestalt structure (e.g., the ‘roof’ evokes the ‘house’). In gesture and whole-body enactments, internal metonymy may motivate processes of profiling and highlighting prototypically essential, or locally salient, aspects of a given action, object, or event. A gesturally

enacted motion onset or manner of motion may evoke the corresponding, fully articulated physical actions (e.g., Figures 2 & 3). Iconic gestures may also portray contours, geometric shapes, spatial dimensions and other relevant qualities of objects, spaces, and other kinds of physical structures. Such internal relations can also be said to link parts and wholes in abstract structures such as schemas, frames, or constructions (e.g., the sentence string in Figure 4a; for more examples see Mittelberg, 2006).

Outer contiguity underlies metonymic expressions in which the profiled element is not part of, but externally contiguous and/or pragmatically related to the element it causes an addressee to infer. For example, the boat in the NYC harbor scene (Figure 1) may evoke the captain and crew. *External metonymy* may draw on various kinds of outer contiguity relations and imply different degrees of metonymic proximity, such as contact, adjacency, impact, and cause/effect (Jakobson & Pomorska, 1983, p. 134). For instance, in “The Élysée Palace remained silent”, the ‘Élysée Palace’ refers to the French President or his spokesperson. The relevant contiguity relations between the building or institution and its inhabitants or members are spatial and pragmatic in nature. House and people belong to the same frame. When people gesture, their hands often create containers or chunks of space for imagined entities which in turn evoke the things talked about and “shown” to interlocutors (Hassemer, 2016; Kendon, 2004; Müller, 2004; Streeck, 2009, 2013). In particular, contact, adjacency, and impact create outer contiguity relations between hands and the objects, tools, and surfaces that speakers are in touch with when communicating. Gestures may highlight, establish, or delete them through metonymic modes operating on them. For example, in Figure 5, the purpose of the gestural enactment is not to iconically imitate someone who is holding something. While this is the perceivable starting point of the meaning construction, it is the imaginary object, immediately contiguous to the two hands seemingly holding it, that the speaker is talking about and is drawing attention to through her speech. Therefore, due to the embodied basic physical action frame of holding an object, it is easy to metonymically infer some middle-sized object, here metaphorically construed.⁷ In the following section, we will look more closely at processes of inferencing that take place between gesturing hands and what they evoke.

7. See Mittelberg & Waugh (2014) for a continuum spanning different degrees and types of body-centered “metonymic proximity” (Jakobson & Pomorska, 1983, p. 134), internal and external metonymy, metonymic shifts and chains, and the interaction of metonymy and metaphor.

Crossmodal pragmatic inferencing and bodily reference points

As suggested throughout this paper, the pragmatically conditioned creation and interpretation of gestures often rely on metonymic principles operating within basic and more complex frames. In this final section, the different threads developed in the course of the paper will be woven together by introducing the notion of *bodily reference points* and taking into account how modality-specific inferential mechanisms play a central role in understanding the motivated and pragmatic aspects of multimodal discourse.

From reference to pragmatic inferencing in multimodal discourse

Incorporating Langacker's (1993) notion of *reference-point constructions*,⁸ Barcelona (2003) stresses the inferential role of metonymies: "conceptual metonymies often provide 'ready-made' pointers towards plausible referential pathways" (p. 97) in interpretative processes. "These pointers, which are normally automatic, contribute greatly to the ease and speed of interpretation" (*ibid.*). Taking these insights, mainly stemming from linguistic investigations, into the realm of bodily semiotics and frame semantics, the point I wish to make is that, in crossmodal processes of pragmatic inferencing – involving speech, gesturing hands, and the invisible things they invoke – metonymic pathways originating from the human body may serve as "natural inference schemata" (Panther & Thornburg, 2003, p. 8) by intuitively drawing on situated ways of being and functioning not only in the physical world, but also in imagined and abstract worlds (e.g., Sweetser, 2007, 2012, on gestures and mental spaces).

... conceptual metonymies such as PART-WHOLE, CAUSE-EFFECT, PERSON-ROLE, REPRESENTATION-REPRESENTED, which have been dubbed *vital relations* by Fauconnier and Turner (2002, pp. 93ff.), are concrete enough to serve as reasoning principles in utterance interpretation. We regard such metonymic relations as *multipurpose conceptual devices* not restricted to language but used in other semiotic systems and thought as well.

(Panther & Thornburg, 2004, pp. 94–95; italics in the original)

Depending on how the concurrent speech cues guide our attention, the focus may stay on the body itself and its gestural action, or it may shift across outer contiguity relations onto the virtual objects that the gesturing hands may be creating or manipulating, or to the locations, discourse contents or people that they may be

8. Langacker (2009) defines the ability of humans to create reference points as follows: "This is our ability to invoke one conceived entity as a reference point in order to establish mental contact with another, i.e. to mentally access one conceived entity through another" (p. 52).

pointing at. Zooming in on the notion of *reference points* (Langacker, 1993, 2009), I propose the term *bodily reference points* to refer to whole-body enactments and gestures that point in more or less subtle ways to something other than themselves while they are constituting physically salient parts of multimodal utterances (Mittelberg, 2006, 2013a). Although prototypical pointing gestures may count as such gestural reference points, what I chiefly have in mind here are gestural indices with lesser degrees of indexicality (see Mittelberg & Waugh, 2014, for a typology of distinct kinds of gestural indices). For example, in certain discourse contexts, visible manual articulators may serve as bodily reference points that trigger access not only to immediately contiguous (virtual) entities, such as implied objects, tools, chunks of space, or lines drawn in the air (Figures 3, 4, 6), but also to less object-like entities or intangible aspects and qualities. These observations further support the idea that in metonymic mappings the source meaning, for instance a gesturing hand and/or movement, remains present and perceptually salient, while the target meaning, for instance the discourse contents (such as the properties of an infinitive; Figures 5a–c) is conceptually prominent in the ongoing exposition (Panther & Thornburg, 2004, p. 95, p. 105).

This discussion brings up the issue of reference, which is one of metonymy's main attested functions (e.g., Panther & Radden, 1999). As compared to language-based accounts, the question of reference poses itself differently when it comes to gesture. Since gestures do not represent an equally conventionalized semiotic system as spoken and signed languages, getting at the meaning of many gestures may turn out to be more a matter of inference than of reference. As McNeill (2005) pointed out, gestures often are what they are taken to be about (see also Mittelberg, 2013a/b). In a similar vein, given the focus on gesture, the idea of cognitive contiguity and cognitive access, stressed in the cognitive linguistics literature (see also Section 4), can only be seen as inseparable from situated, experientially grounded aspects that allow us to mentally enact or sense the gestural forms and actions performed by others. Due to the high degree of cognitive, physical, and pragmatic routinization, addressees may even anticipate the goal-orientedness of certain gestural actions (Clark, 2003).⁹ Habitual metonymic correlations seem to pave the way for such body-based and experientially entrenched inferences (see also Dancygier & Sweetser, 2014, p. 144):

Metonymy has this inferential role because of its ability to mentally activate the implicit pre-existing connection of a certain element of knowledge or experience to another. The referential function of metonymies is thus a useful (hence extremely frequent) consequence of their inference-guiding role since what we do

9. See Rizzolazzi and Arbib (1998) on the understanding of purposeful actions of others, mirror neurons, and language.

when we understand a referential metonymy is to infer the referential intentions of others (Nerlich & Clarke, 2001).
 (Barcelona, 2009, p. 369)

To arrive at the meaning of a multimodal utterance, the addressee has to combine several inferential pathways provided in the different modalities. First, the gesture's full shape or action pattern needs to be derived, at least approximately, from the salient aspects of a hand shape and/or movement, which typically allude rather furtively and schematically to full action routines and what these may entail (e.g., Arnheim, 1969). At the same time, such highly polysemous and potentially multifunctional gestural forms usually receive their local meaning from the speech content and other contextual factors (e.g., Calbris, 2011; Müller, 2004; Streeck et al., 2011). Still, bodily reference points may enhance access to conceptual targets that are not automatically identifiable or easily understandable. Staying within the gesture modality, in the case of the "string of words" (Figures 5a/b), one first needs to shift from the gesturing hands to the imaginative line that they are tracing in the air, and then from there to the abstract concepts and relations that they are metaphorically representing (Mittelberg, 2006). Resulting from a semiotic grounding practice (as described in the section on complex frame structures above), the gestural trace and subsequent grouping actions provide visuo-dynamic support structures that facilitate, as a sort of bodily reference-point construction, accessing the SENTENCE frame and inferring other correlated frames (e.g., the CONSTITUENT frame).¹⁰

Muted degrees of iconicity/indexicality and resulting pragmatic functions

In addition to prototypical pointing gestures and object-oriented gestures, there are other gestures with more muted degrees of indexicality that seem particularly pertinent to gesture pragmatics. When accounting for varying degrees of abstraction from the original, situated groundedness of physical actions, the question of whether we can presuppose some object-like entity as being part of a gesture is not trivial. For example, certain palm-up open hand gestures fulfill predominant discourse functions (Müller, 1998, 2004) or may simply point out the existence of ideas or the relation between ideas (Mittelberg, 2008, 2014, 2017). As shown by Bavelas and colleagues (1992, 1995), interactive gestures may fulfill an array of different functions, such as citing or referring to interlocutors. These practices also have some metonymic import because such *interactive discourse indices*

10. Similar body-based, reference-point mechanisms have been described as a "cognitive key" in understanding certain ASL signs and constructions (P. Wilcox, 2004). Talmi's (2013) reflections on 'gestures as cues to a target' provide additional theoretical foundations for these observations (see also Cienki, 2013a).

(Mittelberg & Waugh, 2014) may lead the interpreting mind to infer the intended meaning present in the semiotic and situational context, for example, from something an interlocutor did or said previously, or from an emotional response he or she displayed while listening. Common ground (Clark, 1996), that is, shared knowledge, experiences, and thus shared frames, can also be pointed at in this fashion (Priesters, 2012; Mittelberg & Rekittke, in press). The ways in which these indexical gestures may be said to partake in multimodal reference-point constructions needs to be examined in more detail across situations and languages.

Of relevance to the present discussion are also those variants of open-palm gestures that employ empty hands to signal obviousness, carelessness, or cluelessness, and thus not perform referential but rather epistemic or modal, that is, predominantly pragmatic functions (Calbris, 1990, 2011; Kendon, 2004, 2017; Müller, 1998, 2004, 2017). As discussed earlier, the palm-up open hand gesture in Figure 5c combines, to varying degrees, referential and pragmatic functions: while one can assume some virtual object metaphorically signifying the infinitive, the gesture also exhibits an interactive – or conative (Jakobson, 1960) – function (requesting an answer from the students) and a modal – emotive (Jakobson, 1960) – function (expressing that to her the answer is obvious).¹¹ We can thus formulate the following proposal: The lesser the degree of iconicity and indexicality, the lesser the referential function, or, put differently, the more closely these bimodally achieved, multifunctional processes seem to be situated at the juncture of ritualization, conventionalization, grammaticalization, and gesture pragmatics.

Conclusion and outlook

The frame-based perspective on gesture pragmatics laid out in this paper is anchored in the realization that speaker-gesturers tend to engage in *doing* something when they are talking, thinking, and interacting, as a means to advance the ongoing discourse and the addressee's understanding of what they are trying to convey. It was argued that such pragmatically minded, multimodal semiotic acts are, at least partly, achieved by linguistically and gesturally foregrounding locally relevant frame elements that may activate – both in the gesturer and in the addressee – larger, dynamically evolving frame structures, metonymic inferences, and metaphor systems. Thus, when gesturing with their hands, heads, or whole bodies, speakers seem to not only simulate physical actions, semiotic practices, and other people's

¹¹. See also Debras (2014) and Schoonjans (2014) on multimodal expressions of stance, Tabacaru (2014) on multimodally achieved humor, and Mittelberg & Waugh (2014) on *modal indices* and *mental or emotional state indices*.

gestures, but to also “do metaphors” (Gibbs, 2017), image schemas, force dynamics (Talmy, 1988), metonymies, and frames (see also, e.g., Mittelberg, 2013a/b, on the *exbodyed mind*; Tabacaru, 2014; Wehling, 2010, 2017); and they do so by adopting shifting viewpoints and moving between different, interrelated mental spaces (Sweetser, 2007, 2012). In concert with the concurrent speech, eye-gaze, and facial expressions, these metonymic gestural actions become integrated into dynamic, meaningful *contextures of action* (Goodwin, 2011), often co-constructed between interlocutors (Müller, 1998; Streeck, 2013; Streeck et al., 2011).

An essential aim of this paper has been to show that, when investigating how gestures participate in, and thus shape, such pragmatically conditioned contexts, it seems preferable to move from predominantly cognitive accounts of metonymy to a truly embodied, body-centered account of frame evocation and metonymy-guided inferences that takes the specific mediality of gestures into consideration. Anchored in the Fillmorian notion of scene, the present proposal hopes to contribute to this endeavor by presenting different kinds of more and less embodied frame structures as well as distinct gestural practices of gestural frame evocation that may trigger processes of pragmatic inferencing. Narrowing in on body-based reference-point constructions (Langacker, 1993) further allows us to revisit the notion of reference and to weigh the varying degrees of indexicality, iconicity, and conventionality in a given gesture. As suggested in the final section of this paper, gestures exhibiting a comparably low degree of iconicity and/or indexicality are likely to assume pragmatic rather than referential functions. One interesting avenue of future research would thus be to examine how pragmatic functions interact with increased levels of abstraction, schematicity, and conventionalization in different languages, including sign languages, and across various discourse genres.

This leads us to a more general area of further inquiry that situates itself at the interface of gesture pragmatics and grammaticalization. Metonymy, metaphor, image schemas, and frames are among the construal operations that generally feed into increasing degrees of schematicity, conventionalization, and grammaticalization in language (e.g., Fillmore, 1988; Sweetser, 1990). To further elaborate and test the theoretical account laid out in this paper, the different tendencies in gestural evocation of scenes and frames proposed here lend themselves to investigate the following assumptions: (1) Through routinized body-based metonymic inferences and metaphoric processes, basic physical action and object frames may build the basis for processes of lexicalization, grammaticalization and simple constructions in gesture; (2) More complex, highly abstract frame structures tend to also underpin and motivate highly schematic constructs exhibiting an internal, linear or hierarchical, structure, such as calendars, films, discourses, theories, as well as certain (multimodal) constructions. These tendencies are assumed to interact and

engender transient cases. To better understand the emergence of such pragmatically conditioned patterns, gesture studies can no doubt greatly benefit from taking into account principles and crossmodal correlations found across a variety of spoken and signed languages.¹²

Acknowledgements

The author wishes to thank the editors for including this article in the special issue on gesture pragmatics, as well as Mary M. Copple, Jennifer Hinnell, and Alexander Ziem for valuable feedback on earlier versions of the manuscript and Yoriko Dixon for providing the gesture drawings. The preparation of this article was supported by the Excellence Initiative of the German Federal and State Governments during a research stay at the Excellence Cluster TOPOI, Humboldt-Universität Berlin, during the spring of 2015.

References

- Andrén, Mats (2010). *Children's gestures from 18 to 30 months*. Lund: Centre for Languages and Literatures, Lund University.
- Andrén, Mats (2014). Multimodal constructions in children: Is the headshake part of language? *Gesture*, 14 (2), 141–170. doi: 10.1075/gest.14.2.02and
- Arnheim, Rudolf (1969). *Visual thinking*. Berkeley & Los Angeles: University of California Press.
- Barcelona, Antonio (2003). The case of a metonymic basis of pragmatic inferencing: Evidence from jokes and funny anecdotes. In Klaus-Uwe Panther & Linda Thornburg (Eds.), *Metonymy and pragmatic inferencing* (pp. 81–102). Amsterdam & Philadelphia: John Benjamins. doi: 10.1075/pbns.113.07bar
- Barcelona, Antonio (2009). Motivation of construction meaning and form: The roles of metonymy and inference. In Klaus-Uwe Panther, Linda L. Thornburg, & Antonio Barcelona (Eds.), *Metonymy and metaphor in grammar* (pp. 363–401). Amsterdam & Philadelphia: John Benjamins. doi: 10.1075/hcp.25.22bar
- Bavelas, Janet, Nicole Chovil, Douglas Lawrie, & Allan Wade (1992). Interactive gestures. *Discourse Processes*, 15, 469–489. doi: 10.1080/01638539209544823
- Bavelas, Janet, Nicole Chovil, Jan L. Coates, & Lori Roe (1995). Gestures specialized for dialogue. *Personality and Social Psychology Bulletin*, 21, 394–405. doi: 10.1177/0146167295214010
- Benazzo, Sandra & Aliyah Morgenstern (2014). A bilingual child's multimodal path into negation. *Gesture*, 14 (2), 171–202. doi: 10.1075/gest.14.2.03ben

¹². For sign language research at the juncture of embodied grammaticalization and multimodal discourse pragmatics see, e.g., Janzen & Shaffer, 2002; Kendon, 2008; Pfau & Steinbach, 2004; S. Wilcox, 2004.

- Benczes, Reka, Antonio Barcelona, & Francisco José Ruiz de Mendoza Ibáñez (Eds.) (2011). *Defining metonymy in cognitive linguistics*. Amsterdam & Philadelphia: John Benjamins. doi: 10.1075/hcp.28
- Bressem, Jana (2012). Repetitions in gesture: Structures, functions, and cognitive aspects. Ph.D. dissertation, European University Viadrina, Frankfurt (Oder).
- Bressem, Jana & Cornelia Müller (2014). The family of AWAY-gestures. In: Cornelia Müller, Alan Cienki, Ellen Fricke, Silvia H. Ladewig, David McNeill, & Jana Bressem (Eds.), *Body – language – communication: An international handbook on multimodality in human interaction*, Volume 2 (pp. 1592–1604). Berlin & Boston: De Gruyter Mouton.
- Calbris, Geneviève (1990). *The semiotics of French gestures*. Cambridge: Cambridge University Press.
- Calbris, Geneviève (2003). From cutting an object to a clear cut analysis. Gesture as the representation of a preconceptual schema linking concrete actions to abstract notions. *Gesture*, 3 (1), 19–46. doi: 10.1075/gest.3.1.03cal
- Calbris, Geneviève (2011). *Elements of meaning in gesture*. Amsterdam & Philadelphia: John Benjamins. doi: 10.1075/gs.5
- Cienki, Alan (1998). Metaphoric gestures and some of their relations to verbal metaphoric expressions. In John-Pierre Koenig (Ed.), *Discourse and cognition: Bridging the gap* (pp. 189–204). Stanford: CSLI Publications.
- Cienki, Alan (2013a). Cognitive Linguistics: Spoken language and gesture as expressions of conceptualization. In Cornelia Müller, Alan Cienki, Elan Fricke, Silvia H. Ladewig, David McNeill, & Sedinha Tessendorf (Eds.), *Body – language – communication: An international handbook on multimodality in human interaction*, Volume 1 (pp. 182–201). Berlin & Boston: Mouton de Gruyter.
- Cienki, Alan (2013b). Mimetic schemas and image schemas in cognitive linguistics and gesture studies. *Review of Cognitive Linguistics*, 11 (2), 417–432. doi: 10.1075/rcl.11.2.13cie
- Cienki, Alan & Cornelia Müller (Eds.) (2008). *Metaphor and gesture*. Amsterdam & Philadelphia: John Benjamins. doi: 10.1075/gs.3
- Clark, Herbert H. (1996). *Using language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620539
- Clark, Herbert H. (2003). Pointing and placing. In Sotaro Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 243–268). Mahwah, N.J. & London: Lawrence Erlbaum Associates.
- Coulson, Seana (2001). *Semantic leaps: Frame-shifting and conceptual blending in meaning construction*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511551352
- Croft, William (1993). The role of domains in the interpretation of metaphor and metonymy. *Cognitive Linguistics*, 4, 335–370. doi: 10.1515/cogl.1993.4.4.335
- Croft, William & Alan A. Cruse (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511803864
- Dancygier, Barbara & Eve E. Sweetser (2005). *Mental spaces in grammar: Conditional constructions*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511486760
- Dancygier, Barbara & Eve E. Sweetser (2012). *Viewpoint in language: A multimodal perspective*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139084727
- Dancygier, Barbara & Eve E. Sweetser (2014). *Figurative language*. Cambridge: Cambridge University Press.

- Debras, Camille (2014). L'Expression multimodale du positionnement interactionnel (multimodal stance-taking). Unpublished doctoral dissertation, Université Sorbonne Nouvelle – Paris 3.
- Dirven, René & Ralf Pörings (Eds.) (2002). *Metaphor and metonymy in comparison and contrast*. Berlin & New York: Mouton de Gruyter.
- Enfield, Nick J. (2011). Elements of formulation. In Jürgen Streeck, Charles Goodwin, & Curtis LeBaron (Eds.), *Embodied interaction: Language and the body in the material world* (pp. 59–66). Cambridge: Cambridge University Press.
- Fauconnier, Gilles & Mark Turner (2002). *The way we think*. New York: Basic Books.
- Feyaerts, Kurt (1999). Metonymic hierarchies. The conceptualization of stupidity in German idiomatic expressions. In: Klaus-Uwe Panther & Günther Radden (Eds.), *Metonymy in language and thought* (pp. 309–332). Amsterdam & Philadelphia: Benjamins.
doi: 10.1075/hcp.4.18fey
- Fillmore, Charles J. (1975). An alternative to checklist theories of meaning. In Cathy Cogen, Henry Thompson, Graham Thurgood, Kenneth Whistler, & James Wright (Eds.), *Proceedings of the first annual meeting of the Berkeley Linguistics Society* (pp. 123–131). Berkeley, CA: Berkeley Linguistics Society.
- Fillmore, Charles J. (1977). Scenes-and-frames semantics. In Antonio Zampolli (Ed.), *Linguistic structures processing*, Volume 4 (pp. 55–81). Amsterdam & New York: North Holland.
- Fillmore, Charles J. (1982). Frame semantics. In *Linguistic Society of Korea* (Ed.), *Linguistics in the morning calm* (pp. 111–137). Seoul: Hanshin.
- Fillmore, Charles J. (1985). Frames and the semantics of understanding. *Quarderni di Semantica*, 6, 222–254.
- Fillmore, Charles J. (1988). The mechanisms of construction grammar. *Berkeley Linguistics Society*, 14, 35–55. doi: 10.3765/bls.v14i0.1794
- Fillmore, Charles J. & Collin F. Baker (2009). A frames approach to semantic analysis. In Bernd Heine & Heiko Narrog (Eds.), *The Oxford handbook of linguistic analysis* (pp. 313–340). Oxford: Oxford University Press.
- Fricke, Ellen (2012). *Grammatik multimodal: Wie Wörter und Gesten zusammenwirken*. Berlin & New York: Mouton de Gruyter. doi: 10.1515/9783110218893
- Gibbs, Jr., Raymond W. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge: Cambridge University Press.
- Gibbs, Jr., Raymond W. (1999). Speaking and thinking with metonymy. In Günter Radden & Zoltan Kövecses (Eds.) *Metonymy in language and thought* (pp. 61–76). Amsterdam & Philadelphia: John Benjamins. doi: 10.1075/hcp.4.04gib
- Gibbs, Jr., Raymond W. (2006). *Embodiment and cognitive science*. Cambridge: Cambridge University Press.
- Gibbs, Jr., Raymond W. (2017). The embodied and discourse view of metaphor: Why these are not so different and how they can be brought closer together. In Beate Hampe (Ed.), *Metaphor: Embodied cognition and discourse* (pp. 319–334). Cambridge: Cambridge University Press.
- Goldberg, Adele (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: The University of Chicago Press.
- Goodwin, Charles (2007). Environmentally coupled gestures. In Susan Duncan, Justine Cassell, & Elena Levy (Eds.), *Gesture and the dynamic dimensions of language* (pp. 195–212). Amsterdam & Philadelphia: John Benjamins. doi: 10.1075/gs.1.18goo

- Goodwin, Charles (2011). Contextures of action. In Jürgen Streeck, Charles Goodwin, & Curtis LeBaron (Eds.), *Embodied interaction: Language and body in the material world* (pp. 182–193). Cambridge: Cambridge University Press.
- Grady, Joe (1997). *Foundations of meaning: primary metaphors and primary scenes*. Doctoral dissertation, University of California at Berkeley.
- Grady, Joe & Christopher Johnson (2002). Converging evidence for the notions of subscene and primary scene. In René Dirven & Ralph Pörings (Eds.), *Metaphor and metonymy in comparison and contrast* (pp. 533–554). Berlin: Mouton de Gruyter.
doi: 10.1515/9783110219197.533
- Grandhi, Sukeshini A., Gina Joue, & Irene Mittelberg (2011). Understanding naturalness and intuitiveness in gesture production: Insights for touchless gestural interfaces. *Proceedings of the ACM2011 conference on Human Factors in Computing Systems (CHI)*, Vancouver, B.C.
doi: 10.1145/1978942.1979061
- Hassemer, Julius (2016). Towards a theory of Gesture Form Analysis: Imaginary forms as part of gesture conceptualisation with empirical support from motion-capture data. Doctoral dissertation, RWTH Aachen University.
- Hopper, Paul & Sandra Thompson (1980). Transitivity in grammar and discourse. *Language*, 56, 251–299.
- Hostetter, Autumn B., & Martha W. Alibali (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, 15 (3), 495–514. doi: 10.3758/PBR.15.3.495
- Jakobson, Roman (1956/1990). Two aspects of language and two types of aphasic disturbances. In Linda R. Waugh & Monique Monville-Burston (Eds.), *Roman Jakobson, On language* (pp. 115–133). Cambridge, MA: Harvard University Press.
- Jakobson, Roman (1960/1987). Linguistics and poetics. In Krystyna Pomorska & Stephen Rudy (Eds.), *Roman Jakobson – Language in literature* (pp. 62–94). Cambridge, MA: Belknap of Harvard University Press.
- Jakobson, Roman & Krystyna Pomorska (1983). *Dialogues*. Cambridge: Harvard University Press.
- Janzen, Terry & Barbara Shaffer (2002). Gesture as substrate in the process of ASL grammaticalization. In Richard P. Meier, Kearsy Cormier, & David Quintos-Pozos (Eds.), *Modality and structure in signed and spoken language* (pp. 199–223). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511486777.010
- Johnson, Mark (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.
- Johnson, Mark (2005). The philosophical significance of image schemas. In Beate Hampe (Ed.), *From perception to meaning: Image schemas in cognitive linguistics* (pp. 15–33). Berlin: Mouton de Gruyter. doi: 10.1515/9783110197532.1.15
- Johnson, Mark (2007). *The meaning of the body: Aesthetics of human understanding*. Chicago: The University of Chicago Press. doi: 10.7208/chicago/9780226026992.001.0001
- Kendon, Adam (1995). Gestures as illocutionary and discourse structure markers in southern Italian conversation. *Journal of Pragmatics*, 23, 247–279. doi: 10.1016/0378-2166(94)00037-F
- Kendon, Adam (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511807572
- Kendon, Adam (2008). Some reflections on the relationship between ‘gesture’ and ‘sign’. *Gesture*, 8 (3), 348–366. doi: 10.1075/gest.8.3.05ken

- Kendon, Adam (2017). Pragmatic functions of gestures: Some observations on the history of their study and their nature. *Gesture*, 16 (2), 157–175 ([this issue](#)).
doi: [10.1075/gest.16.2.01ken](https://doi.org/10.1075/gest.16.2.01ken)
- Labov, William (1972). The transformation of reality in narrative syntax. In William Labov, *Language in the inner city* (pp. 354–396). Philadelphia: The University of Pennsylvania Press.
- Ladewig, Silva H. (2014). Recurrent gestures. In: Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill, & Jana Bressem (Eds.), *Body – language – communication: An international handbook on multimodality in human interaction*, Volume 2 (pp. 1558–1575), Berlin & Boston: Mouton de Gruyter.
- Lakoff, George (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: The University of Chicago Press. doi: [10.7208/chicago/9780226471013.001.0001](https://doi.org/10.7208/chicago/9780226471013.001.0001)
- Langacker, Ronald W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites*, Volume 1. Stanford: Stanford University Press.
- Langacker, Ronald W. (1993). Reference-point constructions. *Cognitive Linguistics*, 4, 1–38.
doi: [10.1515/cogl.1993.4.1.1](https://doi.org/10.1515/cogl.1993.4.1.1)
- Langacker, Ronald W. (2009). Metonymic grammar. In Klaus-Uwe Panther, Linda L. Thornburg, Antonio Barcelona (Eds.), *Metonymy and metaphor in grammar* (pp. 45–71). Amsterdam & Philadelphia: John Benjamins. doi: [10.1075/hcp.25.04lan](https://doi.org/10.1075/hcp.25.04lan)
- Littlemore, Jeanette (2015). *Metonymy: Hidden shortcuts in language, thought and communication*. Cambridge: Cambridge University Press. doi: [10.1017/CBO9781107338814](https://doi.org/10.1017/CBO9781107338814)
- McNeill, David (1992). *Hand and mind: What gestures reveal about thought*. Chicago: The University of Chicago Press.
- McNeill, David (2005). *Gesture and thought*. Chicago: The University of Chicago Press.
doi: [10.7208/chicago/9780226514642.001.0001](https://doi.org/10.7208/chicago/9780226514642.001.0001)
- Mittelberg, Irene (2006). *Metaphor and metonymy in language and gesture: Discourse evidence for multimodal models of grammar* (Doctoral dissertation, Cornell University). Ann Arbor, MI: UMI.
- Mittelberg, Irene (2008). Peircean semiotics meets conceptual metaphor: Iconic modes in gestural representations of grammar. In Alan Cienki & Cornelia Müller (Eds.), *Metaphor and gesture* (pp. 115–154). Amsterdam & Philadelphia: John Benjamins.
- Mittelberg, Irene (2010a). Geometric and image-schematic patterns in gesture space. In Vyvyan Evans & Paul Chilton (Eds.), *Language, cognition, and space: The state of the art and new directions* (pp. 351–385). London: Equinox.
- Mittelberg, Irene (2010b). Interne und externe Metonymie: Jakobsonsche Kontiguitätsbeziehungen in redebegleitenden Gesten. *Sprache und Literatur*, 41 (1), 112–143.
- Mittelberg, Irene (2013a). The exbodyed mind: Cognitive-semiotic principles as motivating forces in gesture. In Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill, & Sedinha Tessendorf (Eds.), *Body – language – communication: An international handbook on multimodality in human interaction* (pp. 750–779). Berlin & Boston: Mouton de Gruyter.
- Mittelberg, Irene (2013b). Balancing acts: Image schemas and force dynamics as experiential essence in pictures by Paul Klee and their gestural enactments. In Michael Bokrent, Barbara Dancygier, & Jennifer Hinnell (Eds.), *Language and the creative mind* (pp. 325–346). Stanford: CSLI.
- Mittelberg, Irene (2014). Gestures and iconicity. In Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill, & Jana Bressem (Eds.), *Body – language – communication*:

- An international handbook on multimodality in human interaction* (pp. 1712–1732). Berlin & Boston: Mouton de Gruyter.
- Mittelberg, Irene (2017). Multimodal existential constructions in German: Manual actions as experiential substrate for grammatical and gestural patterns. *Linguistics Vanguard*, 3, s1. doi: 10.1515/lingvan-2016-0047.
- Mittelberg, Irene & Linda R. Waugh (2009). Metonymy first, metaphor second: A cognitive-semiotic approach to multimodal figures of thought in co-speech gesture. In Charles Forceville & Eduardo Urios-Aparisi (Eds.), *Multimodal metaphor* (pp. 329–356). Berlin & New York: Mouton de Gruyter.
- Mittelberg, Irene & Linda R. Waugh (2014). Gestures and metonymy. In: Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill, & Jana Bressem (Eds.), *Body – language – communication: An international handbook on multimodality in human interaction* (pp. 1747–1766). Berlin & Boston: Mouton de Gruyter.
- Mittelberg, Irene & Gina Joue (2017). Source actions ground metaphor via metonymy: Toward a frame-based account of gestural action in multimodal discourse. In Beate Hampe (Ed.), *Metaphor: Embodied cognition and discourse* (pp. 119–137). Cambridge: Cambridge University Press.
- Müller, Cornelia (1998). *Redebegleitende Gesten: Kulturgeschichte – Theorie – Sprachvergleich*. Berlin: Spitz Verlag.
- Müller, Cornelia (2004). Forms and uses of the palm up open hand: A case of a gesture family? In Cornelia Müller & Roland Posner (Eds.), *The semantics and pragmatics of everyday gesture: The Berlin conference* (pp. 233–256). Berlin: Weidler Verlag.
- Müller, Cornelia (2017). How recurrent gestures mean: Conventionalized contexts-of-use and embodied motivation. *Gesture*, 16 (2), 277–304 (this issue).
- Nerlich, Brigitte & David D. Clarke (2001). Ambiguities we live by: Towards a pragmatics of polysemy. *Journal of Pragmatics*, 33 (1), 1–20. doi: 10.1016/S0378-2166(99)00132-0
- Núñez, Rafael E. (2008). A fresh look at the foundations of mathematics: Gesture and the psychological reality of conceptual metaphor. In Alan Cienki & Cornelia Müller (Eds.), *Metaphor and gesture* (pp. 93–114). Amsterdam & Philadelphia: John Benjamins.
- Panther, Klaus-Uwe & Günter Radden (Eds.) (1999). *Metonymy in language and thought*. Amsterdam & Philadelphia: John Benjamins. doi: 10.1075/hcp.4
- Panther, Klaus-Uwe & Linda L. Thornburg (Eds.) (2003). *Metonymy and pragmatic inferencing*. Amsterdam & Philadelphia: John Benjamins. doi: 10.1075/pbns.113
- Panther, Klaus-Uwe & Linda L. Thornburg (2004). The role of conceptual metonymy in meaning construction. *Metaphorik.de*, 06, 91–113.
- Panther, Klaus-Uwe, Linda L. Thornburg, & Antonio Barcelona (Eds.) (2009). *Metonymy and metaphor in grammar*. Amsterdam & Philadelphia: John Benjamins Publishing. doi: 10.1075/hcp.25
- Parrill, Fey (2009). Dual viewpoint in gesture. *Gesture*, 9 (3), 271–289. doi: 10.1075/gest.9.3.01par
- Parrill, Fey & Eve E. Sweetser (2004). What we mean by meaning: Conceptual integration in gesture analysis and transcription. *Gesture*, 4 (2), 197–219. doi: 10.1075/gest.4.2.05par
- Peirce, Charles S. (1960). *Collected papers of Charles Sanders Peirce (1931–1958)*, Volume I: *Principles of philosophy*; Volume II: *Elements of logic*. Charles Hartshorne & Paul Weiss (Eds.). Cambridge, MA: The Belknap of Harvard University Press.

- Pfau, Roland & Markus Steinbach (2004). On grammaticalization: Do sign languages follow the well-trodden paths? *Proceedings of the 8th Theoretical Issues in Sign Language Research (TISLR 8) Conference, Barcelona*.
- Peirsman, Yves & Dirk Geeraerts (2006). Metonymy as a prototypical category. *Cognitive Linguistics*, 17 (3), 269–316. doi: 10.1515/COG.2006.007
- Priesters, Matthias A. (2012). *Functional patterns in gesture space. Correlating location and function of gestures using 3D motion capture technology*. Unpublished M.A. thesis, RWTH Aachen University.
- Rizzolazzi, Giacomo & Michael A. Arbib (1998). Language within our grasp. *Trends in Neurosciences*, 21 (5), 188–194.
- Schank, Robert. C. & Robert Abelson (1977). *Scripts, goals, plans, and understanding*. Hillsdale, NJ: Lawrence Erlbaum.
- Schoonjans, Steven (2014). *Modalpartikeln als multimodale Konstruktionen. Eine korpusbasierte Kookkurrenzanalyse von Modalpartikeln und Gestik im Deutschen*. Unpublished doctoral dissertation, KU Leuven.
- Schoonjans, Steven, Paul Sambre, Geert Brône, & Kurt Feyaerts (2016). Vers une analyse multimodale du sens. Perspectives constructionnelles sur la gestualité co-grammaticale. *Languages*, 201, 33–50.
- Slobin, Dan (1985). Crosslinguistic evidence for the language-making capacity. In Dan Slobin (Ed.), *A crosslinguistic study of language acquisition. Volume 2: Theoretical issues* (pp. 1157–1256). Hillsdale, NJ: Lawrence Erlbaum.
- Sperber, Dan & Deirdre Wilson (1986/1995). *Relevance: Communication and cognition* (2nd Ed.), Oxford: Blackwell.
- Steen, Francis & Mark Turner (2013). Multimodal construction grammar. In Mike Borkent, Barbara Dancygier, & Jennifer Hinnell (Eds.), *Language and the creative mind* (pp. 255–274), Stanford: CSLI.
- Streeck, Jürgen (2009). *Gesturecraft: The Manu-facture of meaning*. Amsterdam & Philadelphia: John Benjamins. doi: 10.1075/gs.2
- Streeck, Jürgen (2013). Praxeology of gesture, In Cornelia Müller, Alan Cienki, Ellen Fricke, Silvia H. Ladewig, David McNeill, & Jana Bressem (Eds.), *Body – language – communication. An international handbook of multimodality in human interaction*, Volume 1 (pp. 674–688). Berlin & Boston: Mouton de Gruyter.
- Streeck, Jürgen, Charles Goodwin, & Curtis LeBaron (Eds.) (2011). *Embodied interaction: Language and the body in the material world*. Cambridge: Cambridge University Press.
- Sweetser, Eve E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620904
- Sweetser, Eve E. (1998). Regular metaphoricity in gesture: Bodily-based models of speech interaction. *Actes du 16e Congrès International des Linguistes* (CD-ROM), Elsevier.
- Sweetser, Eve E. (2007). Looking at space to study mental spaces: Co-speech gesture as a crucial data source in cognitive linguistics. In Monica Gonzalez-Marquez, Irene Mittelberg, Seana Coulson, & Michael Spivey (Eds.), *Methods in cognitive linguistics* (pp. 201–224). Amsterdam & Philadelphia: John Benjamins. doi: 10.1075/hcp.18.15swe
- Sweetser, Eve E. (2012). Viewpoint and perspective in language and gesture: From the ground down. In Barbara Dancygier & Eve E. Sweetser (Eds.), *Viewpoint in language: A multimodal perspective* (pp. 1–24). Cambridge: Cambridge University Press.
doi: 10.1017/CBO9781139084727.002

- Sullivan, Karen (2013). *Frames and constructions in metaphoric language*. Amsterdam & Philadelphia: John Benjamins. doi: 10.1075/cal.14
- Tabacaru, Sabina (2014). Humorous implications and meanings: A multimodal study of sarcasm in interactional humor. Unpublished doctoral dissertation, Université Charles de Gaulle – Lille 3.
- Talmy, Leonard (1988). Force dynamics in language and cognition. *Cognitive Science*, 12, 49–100. doi: 10.1207/s15516709cog1201_2
- Talmy, Leonard (2013). Gestures as cues to a target. Manuscript of talk presented at the ICLC 12, University of Edmonton, Alberta, June 24, 2013.
- Waugh, Linda R. & Monique Monville-Burston (1990). Roman Jakobson: His life, work and influence. Introduction to: Linda R. Waugh & Monique Monville Burston (Eds.), *Jakobson on language* (pp. 1–45). Cambridge: Belknap of Harvard University Press.
- Wehling, Elisabeth (2010). Argument is gesture war: Function, form and prosody of discourse structuring gestures in political argument. *Proceedings of the 35th Annual Meeting of the Berkeley Linguistics Society* (pp. 54–65). Berkeley, CA: Berkeley Linguistics Society.
- Wehling, Elisabeth (2017). Discourse management gestures. *Gesture*, 16 (2), 245–276 (this issue).
- Wilcox, Phyllis P. (2004). A cognitive key: Metonymic and metaphorical mappings in ASL. *Cognitive Linguistics*, 15 (2), 197–222. doi: 10.1515/cogl.2004.008
- Wilcox, Sherman (2004). Gesture and language: Cross-linguistic and historical data from signed languages. *Gesture*, 4, 43–73. doi: 10.1075/gest.4.1.04wil
- Ziem, Alexander (2014). *Frames of understanding in text and discourse*. Amsterdam & Philadelphia: John Benjamins. doi: 10.1075/hcp.48
- Zima, Elisabeth (2014). Gibt es multimodale Konstruktionen? Eine Studie zu [V(motion) in circles] und [all the way from X PREP Y], *Gesprächsforschung*, 15, 1–48.
- Zima, Elisabeth & Alexander Bergs (Guest Eds.) (2017). *Toward a multimodal construction grammar* (Special Issue). *Linguistics Vanguard*, 3, s1.
- Zlatev, Jordan (2005). What's in a schema? Bodily mimesis and the grounding of language. In Beate Hampe & Joe Grady (Eds.), *From perception to meaning: Image schemas in cognitive linguistics* (pp. 313–342). Berlin & New York: Mouton de Gruyter.
doi: 10.1515/9783110197532.4.313

Author's address

Irene Mittelberg
 RWTH Aachen University
 Human Technology Centre (HumTec)
 Centre for Sign Language and Gesture (SignGes)
 Theaterplatz 14
 52070 Aachen
 Germany
 mittelberg@humtec.rwth-aachen.de

Biographical notes

Irene Mittelberg is professor of linguistics and cognitive semiotics at the Human Technology Centre of RWTH Aachen University, where she directs the Natural Media Lab and the Centre for Sign Language and Gesture (SignGes). After completing an M.A. in French linguistics and art history at Hamburg University, she received her Ph.D. in linguistics and cognitive studies from Cornell University. Her dissertation (under the guidance of Linda Waugh) investigates how embodied metonymy, metaphor, and image schemas jointly motivate gestures in multimodal explanations of grammar and linguistic theory. Further research interests include grounded abstraction in gesture and painting and the use of motion-capture technology in gesture research.