

Metaphor

Embodied Cognition and Discourse

Edited by

Beate Hampe

Universität Erfurt, Germany

CAMBRIDGE
UNIVERSITY PRESS



University Printing House, Cambridge CB2 8BS, United Kingdom
One Liberty Plaza, 20th Floor, New York, NY 10006, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
4843/24, 2nd Floor, Ansari Road, Daryaganj, Delhi – 110002, India
79 Anson Road, #06-04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.
It furthers the University's mission by disseminating knowledge in the pursuit of
education, learning, and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781107198333

DOI: 10.1017/9781108182344

© Cambridge University Press 2017

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2017

Printed in the United Kingdom by Clays, St Ives plc

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloguing-in-Publication Data

Names: Hampe, Beate, editor.

Title: Metaphor : embodied cognition and discourse / edited by Beate Hampe,

Universität Erfurt, Germany.

Description: Cambridge : Cambridge University Press, 2017. | Includes bibliographical
references and index.

Identifiers: LCCN 2016059378 | ISBN 9781107198333 (hardback : alk. paper)

Subjects: LCSH: Cognitive neuroscience. | Metaphor. | Psycholinguistics.

Classification: LCC QP360.5 .M477 2017 | DDC 612.8/233–dc23

LC record available at <https://lccn.loc.gov/2016059378>

ISBN 978-1-107-19833-3 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy
of URLs for external or third-party internet websites referred to in this publication
and does not guarantee that any content on such websites is, or will remain,
accurate or appropriate.

Contents

	page vii
<i>List of Figures</i>	x
<i>List of Tables</i>	xi
<i>Notes on Contributors</i>	xii
<i>Editor's Preface and Acknowledgments</i>	xvii
Introduction	1
1 Embodiment and Discourse: Dimensions and Dynamics of Contemporary Metaphor Theory BEATE HAMPE	3
Part I: Metaphor in Cognition	25
2 Sources and Targets in Primary Metaphor Theory: Looking Back and Thinking Ahead JOSEPH E. GRADY & GIORGIO A. ASCOLI	27
3 The Hierarchical Structure of Mental Metaphors DANIEL CASASANTO	46
4 Metaphorical Directionality: The Role of Language YESHAYAHU SHEN & ROY PORAT	62
5 Body Schema and Body Image in Metaphorical Cognition VALENTINA CUCCIO	82
6 Primary Metaphors Are Both Cultural and Embodied BODO WINTER & TEENIE MATLOCK	99
Part II: More than Metaphor	117
7 Source Actions Ground Metaphor via Metonymy: Toward a Frame-Based Account of Gestural Action in Multimodal Discourse IRENE MITTELBURG & GINA JOUE	119

8 Metaphor and Other Cognitive Operations in Interaction: From Basicity to Complexity	138
FRANCISCO J. RUIZ DE MENDOZA IBÁÑEZ	
9 On the Role of Embodied Cognition in the Understanding and Use of Metonymy	160
JEANNETTE LITTLEMORE	

Part III: Metaphor in Discourse

179	
10 The Cancer Card: Metaphor, Intimacy, and Humor in Online Interactions about the Experience of Cancer	181
ELENA SEMINO & ZSÓFIA DEMJÉN	
11 Mappings and Narrative in Figurative Communication	200
ALICE DEIGNAN	
12 Contextual Activation of Story Simulation in Metaphor Comprehension	220
L. DAVID RITCHIE	
13 From Image Schema to Metaphor in Discourse: The FORCE Schemas in Animation Films	239
CHARLES J. FORCEVILLE	
14 Doing Metaphor: An Ecological Perspective on Metaphoricity in Discourse	257
THOMAS W. JENSEN	
277	
15 Attention to Metaphor: Where Embodied Cognition and Social Interaction Can Meet, but May Not Often Do So	279
GERARD J. STEEN	
16 Waking Metaphors: Embodied Cognition in Multimodal Discourse	297
CORNELIA MÜLLER	
317	
Epilogue (A Personal View)	
17 The Embodied and Discourse Views of Metaphor: Why These Are Not So Different and How They Can Be Brought Closer Together	319
RAYMOND W. GIBBS, JR.	
335	
Person Index	366
Subject Index	369

Figures

page 76	
4.1 A visual hybrid (half man, half tree)	78
4.2 Example of associated entities	79
4.3 Example of associated entities	
6.1 MORE IS UP expressed in co-speech gesture on the phrase “high number.” The hand starts low and moves up to a higher position (a), with the end point shown in (b).	102
6.2 Co-speech gesture expressing SOCIAL DISTANCE IS PHYSICAL DISTANCE. First the hands are close together (a), then farther apart (b).	
6.3 Spatial positions of characters expressing SOCIAL DISTANCE IS PHYSICAL DISTANCE in <i>Before Midnight</i> . (a) Jesse criticizes Celine for arguing when they should be making love instead; (b) Celine retaliates by blaming Jesse for having started it.	105
6.4 Spatial positions of characters expressing SOCIAL DISTANCE IS PHYSICAL DISTANCE in <i>Before Midnight</i> . Jesse (a) and Celine (b) in a friendly interaction about a gift from their friends.	
6.5 Co-speech gesture expressing SOCIAL DISTANCE IS PHYSICAL DISTANCE. Both hands move toward a location in the center of the body for expressing SIMILARITY (a). For expressing DIFFERENCE, the hands move apart (b).	107
7.1 Multimodal instruction about the grammatical category infinitive	123
7.2 Feeling pulled into a film	132
7.3 Explaining constituent structure	133
8.1 Survey of cognitive operations working on frames and image schemas	141
8.2 Single-source amalgam for “My boss is a pig”	153
8.3 Double-source amalgam for “Slap some common sense into someone”	
8.4 Creation of irony in “Mary is an angel”	154
11.1 Life as a journey through time	158
11.2 The Ph.D. journey	204

7 Source Actions Ground Metaphor via Metonymy: Toward a Frame-Based Account of Gestural Action in Multimodal Discourse

Irene Mitzelberg & Gina Joue*

Chapter Preview

This chapter starts from the observation that metaphoric understandings expressed monomodally through gesture tend to rely on “primary metaphors” (Grady 1997a). Asserting that gestures draw on basic, experientially motivated, embodied construal operations, we detail how primary scenes and subscenes (Grady & Johnson 2002), image and force schemas, metonymy, and frames (Fillmore 1982) interact in situated meaning-making. We propose that by shifting the focus from object-oriented schemas, source domains, and mappings to what we call “source actions” and “embodied action frames,” we can account for the pragmatically minded nature and specific mediality of communicative gestural acts integrated in natural multimodal discourse. We argue that verbal gestures recruit frame structures metonymically, singling out elements of “scenes” (Fillmore 1977), especially those underpinning correlated metaphoric meanings. We back up our theoretical claims with evidence from neuroscientific studies and outline a frame-based approach that helps trace avenues for further research into embodied cognition and multimodal discourse processes.

1 Introduction

Our understanding of the world and the discourses we participate in integrates pieces of information presented in different sensory modalities into dynamic meaningful wholes. As embodied semiotic beings, we observe our conversational partners while listening to what they are saying, we sense emotional qualities in their articulation, and we might reach out verbally and gesturally to them or hold ourselves back. Oral communication clearly involves more than language use; it combines various sign systems such as words, gestures, eye gaze, head movements, and body posture. Naturalistic discourse thus typically consists of

concerted *multimodal acts of contextualized meaning-making*, i.e. of dynamic “semiotic contexts” (Jakobson 1956) or “contexts of action” (Goodwin 2011b). As physical actions of the human body influenced by cognitive and affective states, gestures have been shown to provide valuable insights into the physical grounding and sociocultural situatedness of embodied cognition and language use. Since speakers tend to be mostly unaware of their gestural behavior, gestures may also reveal less monitored aspects of cognitive and emotional processes during communication, including mental simulation and metaphorical construal (e.g. Cienki & Müller 2008a; Gibbs 2006a; Müller 1998, *this volume*; Sweetser 1998, 2007).

This chapter puts the human body and its communicative behavior at the very center of studying embodiment, discourse, metaphor, and ad-hoc meaning construction. We bring together several theoretical constructs to help us understand why the communicating body during spoken discourse production has a strong tendency to simulate deeply entrenched physical routines including motor programs and object manipulation, regardless of whether the discourse topic is abstract or concrete. This approach allows us to account for our observation that metaphors expressed monomodally through gesture alone tend to be “primary metaphors” drawing on “primary scenes” (Grady & Johnson 2002) as well as on basic “scenes” and “frames” (Fillmore 1977, 1982).¹

Following the frame-based account of metaphor and metonymy advanced by Dancygier and Sweetser (2014), we will apply the notion of frames to gesture and draw connections to interrelated embodied construal operations. Shifting the focus from object-oriented source domains to the actions of the human body, we introduce the notion of “source action” in metaphorically construed semiotic acts, which may or may not involve (imagined) persons, objects, or surfaces. Our approach is based on the assumption that metaphor in gesture is tightly linked to metonymic modes. In fact, metonymy grants access to meaning, including metaphorical meaning (Mittelberg & Waugh 2009, 2014).

In more general cognitive terms, we propose that metonymy can be seen as a mechanism through which top-down (cognitive) control filters out irrelevant bottom-up (sensory) features perceived during the comprehension of multimodal discourse. In reducing the features requiring attention for further processing, cognitive control via metonymy provides a means of integrating features with each other and establishing associations with existing semantic frames. Without such a mechanism, the numerous sets of semantic frames that are involved in metaphor understanding, especially in multimodal communication, could be computationally intractable.

We illustrate these premises mainly with data from academic discourse and film retellings in American English and German, and also bring in additional evidence from neuroscience studies. Our findings attest to the tight junction of action and perception and of physical and mental processes. They also shed light on people’s speech-independent understanding of metaphorically motivated actions performed by others who are speaking (Gallese & Lakoff 2005). Zeroing in on “the meaning of the body” (Johnson 2007), that is, speakers’ semiotic actions and semanticized gesture spaces as the immediate, natural media and loci of recurring interactions with the material and social world, our intent is to provide gestural support for the idea that people engaged in discourse *do* metaphor (*this volume*: Gibbs; Müller; Jensen).

2 Pragmatic Foundations of Metaphoric Processes in Multimodal Discourse

2.1 Embodiment: Bodily Semiotic Acts Mediate between Experiential Grounding and Conceptual Schematics

Embodiment, gesture, metaphor, and metonymy are intimately linked. Embodiment theorists generally agree that our higher cognitive and linguistic abilities are shaped by the architecture of our bodies and by the ways in which we interact with our physical and social environment (e.g. Gibbs 2006a; Hampe 2005a; Johnson 1987, 2007; Lakoff & Johnson 1999). Obviously constrained by the morphology of the human body, gestures are also shaped by the forms and affordances of our material habitat and by social interactions in both personal and institutional settings (e.g. Kendon 2004). In particular, many of the basic habitual actions of object manipulation, tactile exploration, and sensorimotor routines that humans use to comprehend the world and to function in it manifest themselves in gestural hand movements performing communicative actions (e.g. Calbris 2011; Müller 1998; Streeck 2009; Streeck et al. 2011).

Questions of how multimodal communicative acts may be driven by embodied conceptual structures and construal operations have inspired a growing branch of gesture research within cognitive linguistics (for overviews, see Cienki 2013a; Sweetser 2007). Building on the premises of embodiment, Mittelberg’s (2008, 2013b) notion of the “exbodied mind,” or the mind externalized through the body, captures this perspective; it accounts for how embodied cognitive-semiotic principles (e.g. iconicity, metaphoricity, indexicality, conventionality, image and force schemas, metonymy, and viewpoint) may not only motivate and structure observable patterns of gestural sign constitution but also guide how gestures are interpreted due to the embodied nature of their motivation. Given the above observations, we expect gestures to



Figure 7.1 Multimodal instruction about the grammatical category infinitive

Consider, for example, the following metaphorically construed gestural portrayal of a grammatical category. During a German language course for beginners, the instructor asks her students about the properties of the grammatical category “infinitive” (Figure 7.1a–c): “You are looking at an ... infinitive ... can you say ... ah, this is ... present, perfect, past perfect [tense]?”² In Figure 7.1a, the instructor is enacting what she wants her students to do, namely, to contemplate the grammatical category “infinitive” that she is speaking about (“you are looking at an infinitive”) and seemingly holding in her hands. She subsequently directs her gaze toward her audience (Figure 7.1b) and continues to hold the bimanual gesture while asking the students whether infinitives may be marked for tense (“can you say ... ah, this is ... present tense, perfect tense, simple past?”). While she is waiting for a response, her right hand forms into a Palm-Up Open Hand (PUOH, Müller 2004; Kendon 2004), and her left hand also opens up slightly (Figure 7.1c). She remains silent in this position for about four seconds. This PUOH qualifies as an interactive *seeking gesture* (Bavelas et al. 1995) signaling the speaker’s wish to receive an answer. The open hand may also indicate that the reified infinitive – now construed as a smaller, imagined object sitting on the open palm – is still available for inspection.

This multimodal teaching performance illustrates both the strikingly dynamic, multidimensional character of embodiment (Gibbs 2006a) and the multifunctional nature of gestures mediating metaphoric understandings. Following Jakobson (1960), we assume a hierarchical ordering of semiotic modes and discourse functions in gestural sign processes (Mittelberg 2013b: 755). The metaphoric dimension of the gesture in Figure 7.1a pertains to the *referential* or *cognitive* function by evoking or referring to an idea. Additional subordinated functions are *emotive* (expressing the speaker’s attitude and intentions), *conative* (making the students think and respond), *phatic* (maintaining the connection with the students, keeping their attention focused on the question),

exhibit a natural tendency to evoke the kinds of metaphors, metonymies, frames, and other construal operations that rely on a high degree of embodiment and reflect their inherent indexicality (e.g. Mittelberg & Waugh 2014).

2.2 Multifunctionality of (Primarily Metaphoric) Gestures

Most gesture scholars agree that gestures do not lend themselves to being assigned to a single category (e.g. McNeill 1992, 2005; Müller 1998); hence, the term “metaphoric gesture” does not do justice to this medium’s versatility. While space limitations do not allow for a detailed discussion of gesture functions, the polysemous and multifunctional nature of gestural signs is briefly addressed here, as it feeds directly into the multidimensional fabric of multimodally expressed metaphoricity.

Consider, for example, the following metaphorically construed gestural portrayal of a grammatical category. During a German language course for beginners, the instructor asks her students about the properties of the grammatical category “infinitive” (Figure 7.1a–c): “You are looking at an ... infinitive ... can you say ... ah, this is ... present, perfect, past perfect [tense]?”² In Figure 7.1a, the instructor is enacting what she wants her students to do, namely, to contemplate the grammatical category “infinitive” that she is speaking about (“you are looking at an infinitive”) and seemingly holding in her hands. She subsequently directs her gaze toward her audience (Figure 7.1b) and continues to hold the bimanual gesture while asking the students whether infinitives may be marked for tense (“can you say ... ah, this is ... present tense, perfect tense, simple past?”). While she is waiting for a response, her right hand forms into a Palm-Up Open Hand (PUOH, Müller 2004; Kendon 2004), and her left hand also opens up slightly (Figure 7.1c). She remains silent in this position for about four seconds. This PUOH qualifies as an interactive *seeking gesture* (Bavelas et al. 1995) signaling the speaker’s wish to receive an answer. The open hand may also indicate that the reified infinitive – now construed as a smaller, imagined object sitting on the open palm – is still available for inspection.

This multimodal teaching performance illustrates both the strikingly dynamic, multidimensional character of embodiment (Gibbs 2006a) and the multifunctional nature of gestures mediating metaphoric understandings. Following Jakobson (1960), we assume a hierarchical ordering of semiotic modes and discourse functions in gestural sign processes (Mittelberg 2013b: 755). The metaphoric dimension of the gesture in Figure 7.1a pertains to the *referential* or *cognitive* function by evoking or referring to an idea. Additional subordinated functions are *emotive* (expressing the speaker’s attitude and intentions), *conative* (making the students think and respond), *phatic* (maintaining the connection with the students, keeping their attention focused on the question),

and *meta-linguistic* (defining the technical term “infinitive”). The speaker’s facial expressions and head movements further enhance and contextualize these functions. Importantly, all these facets of action and interaction may also be observed when people actually handle and exchange physical objects. We will revisit this example to highlight how metonymic and metaphoric modes tend to interact in such experientially motivated, action-based processes of multimodal meaning-making.

3 Strongly Embodied Construal Operations Motivating Metaphoric Processes

In multimodal discourse, metaphorical conceptualizations themselves are often multimodally expressed (e.g. Forceville & Urios-Aparisi 2009). The same metaphorical understanding can be concurrently reflected in both speech and gesture. Different metaphors can also manifest themselves in the two modalities. Additionally, metaphorical understandings may be expressed only in the verbal or only in the gestural modality, thus giving rise to *monomodal* metaphoric expressions (e.g. Müller & Cienki 2009). When basic physical actions and/or the implied objects are set in a metaphorical frame through the concurrent speech and other contextual factors, as in the example of the monomodal metaphor discussed above (Figure 7.1a–c), the tight correlation between basic recurrent physical experiences and metaphorical processes becomes literally tangible. In this section we discuss experientially basic construal operations that give rise to bodily expressed metaphoric processes integrated in multimodal discourse.

3.1 Experiential Basics: Primary Metaphors and Scenes

Grady (1997a, see also Grady & Ascoli, *this volume*) coined the term *primary metaphor* for metaphorical mappings that arise through recurrent perceptual

and sensorimotor experiences influencing people's subjective understanding of the physical and social world (e.g. AFFECTION IS WARMTH; IDEAS ARE FOOD; CAUSES ARE FORCES; DIFFICULTIES ARE BURDENS). Such basic metaphors have first-order relations with corresponding, often physical, experiences. They have been found to motivate both linguistic and gestural expressions in spontaneous discourse about, for instance, abstract domains, moral values, emotions, and mental states (e.g. Cienki 1998; Cienki & Müller 2008b; Mittelberg 2008; Müller 2008b, *this volume*; Parrill & Sweetser 2004; Winter & Matlock, *this volume*). In particular, body posture and kinetic action may reflect essential facets of metaphorical understandings of ideas, categories, events, or relations even when the concurrent speech is non-figurative (e.g. Mittelberg 2006).

Other basic construal mechanisms such as primary scenes (or recurrent experiential scenarios) and subsценen (e.g. Grady & Johnson 2002) may also become manifest mono- or multimodally. They provide ways of deconstructing experiential types into their constitutive parts and dimensions. For example, for the primary scene “manipulation of complex object” (*ibid.*: 537), gestural practices may pick out a particular subscene. We can imagine someone pretending to carry a huge monitor or to hold a small brittle vase. As pointed out by Grady and Johnson, primary metaphors and scenes are akin to what Slobin (1985), in his work on the acquisition of grammatical categories, refers to as “prototypical events,” which also include object manipulation and object transfer. Prototypical events “could include grasping, pushing, squeezing, pulling apart, etc., each of which would be a distinct subscene” (Grady & Johnson 2002: 548).

3.2 Varying Levels of Abstraction: Mimetic Schemas, Image Schemas, and Force Gestalts

As for language development, Zlatev's (2005) notion of “mimetic schema” (e.g. KICK, JUMP, GRASP, EAT, HIT, KISS) as well as Andrén's (2010) notion of “action gestalt” and “action-based gestures” capture intermediate levels of abstraction regarding the understanding and communicative use of actions and objects. Mimetic schemas are “dynamic, concrete and preverbal representations, involving the body image, accessible to consciousness and pre-reflectively shared in a community” (Zlatev 2005: 334). An interesting question is how these schemas and actions are used metaphorically (Cienki 2013b). In adults, higher levels of abstraction can be found in certain “gesture families” (e.g. Calbris 2011; Kendon 2004; Müller 2004) and in “recurrent gestures,” which tend to be more conventionalized (e.g. the cyclic gesture, Ladewig 2011).

As embodied patterns of recurrent experiences with the physical and social world, “image schemas” (e.g. Hampe 2005a; Johnson 1987; Lakoff 1987) are more abstract and schematic than the bodily action routines from which they, at least partly, derive. They have been attested a prominent role not only in structuring our bodily experience but also in motivating metaphoric expressions in diverse modalities (Forceville & Urios-Aparisi 2009; Gibbs 2006). A growing body of research suggests that gestures and body postures may also reflect a range of image schemas, such as PATH, CONTAINMENT, CYCLE, OBJECT, and SUPPORT, as well as “force-dynamic” gestalts (Talmy 1988) such as FORCE, BLOCKAGE, and GRAVITY (e.g. Cienki 2005, 2013b; Ladewig 2011; Mittelberg 2006, 2010, 2013b; Sweetser 1998). In fact, there seems to be a certain structural similarity between image/force schemas and gestural forms and movements. While gestures are physical, they also are schematic gestalts of experiential essence: They typically consist of reduced hand configurations, motion onsets or traces that minimally suggest, for instance, a PATH, an OBJECT (Figure 7.1), CONTAINMENT, or an IN-OUT spatial relation (Figure 7.3). Likewise, subtle shifts in body posture may evoke the sensation of physical forces impacting the speaker's body, thus recruiting, for instance, the BALANCE schema in construals of mental states (e.g. Mittelberg 2013a).

Both mimetic schemas and image schemas motivate gestures and connect the meaning of conventionalized linguistic expressions to physical actions (“symbol grounding problem,” Barsalou 1999). They are different, however, in that the image schema framework focuses more on the links between actions and embodied conceptual structures, whereas the focus in mimetic schemas is on the actions themselves, bringing in imitation, or mimesis, to explain how symbolic meanings arise. Both kinds of schemas are thus congruent with Piaget's proposed role of schemas in cognitive development, namely, as basic cognitive units developed through action and perceptuo-motor interaction (Piaget 1967).

3.3 Mental Representations, Modality, and Body Schema

Being inherently physical, verbal gestures provide a link between mental representations of sensory patterns and symbols characteristic of communication, including embodied abstract concepts. With gesture understanding, where visual input arguably needs to be matched with action representations, there is a more direct computational motivation to maintain fidelity to the “body schema” (Blanke 2012; Gallagher 2005), which can be glossed as the dynamic sense of one's own body (for a discussion of “body schema” vs. “body image,” see Cuccio, *this volume*). We saw this fidelity when we asked healthy adults to demonstrate transitive actions while pretending that a part of their body, e.g. their hand, was the tool, when a tool was required (e.g. to slice an apple). They failed to do so 77.5 percent of the time despite repeated

world, “image schemas” (e.g. Hampe 2005a; Johnson 1987; Lakoff 1987) are more abstract and schematic than the bodily action routines from which they, at least partly, derive. They have been attested a prominent role not only in structuring our bodily experience but also in motivating metaphoric expressions in diverse modalities (Forceville & Urios-Aparisi 2009; Gibbs 2006). A growing body of research suggests that gestures and body postures may also reflect a range of image schemas, such as PATH, CONTAINMENT, CYCLE, OBJECT, and SUPPORT, as well as “force-dynamic” gestalts (Talmy 1988) such as FORCE, BLOCKAGE, and GRAVITY (e.g. Cienki 2005, 2013b; Ladewig 2011; Mittelberg 2006, 2010, 2013b; Sweetser 1998). In fact, there seems to be a certain structural similarity between image/force schemas and gestural forms and movements. While gestures are physical, they also are schematic gestalts of experiential essence: They typically consist of reduced hand configurations, motion onsets or traces that minimally suggest, for instance, a PATH, an OBJECT (Figure 7.1), CONTAINMENT, or an IN-OUT spatial relation (Figure 7.3). Likewise, subtle shifts in body posture may evoke the sensation of physical forces impacting the speaker's body, thus recruiting, for instance, the BALANCE schema in construals of mental states (e.g. Mittelberg 2013a).

Both mimetic schemas and image schemas motivate gestures and connect the meaning of conventionalized linguistic expressions to physical actions (“symbol grounding problem,” Barsalou 1999). They are different, however, in that the image schema framework focuses more on the links between actions and embodied conceptual structures, whereas the focus in mimetic schemas is on the actions themselves, bringing in imitation, or mimesis, to explain how symbolic meanings arise. Both kinds of schemas are thus congruent with Piaget's proposed role of schemas in cognitive development, namely, as basic cognitive units developed through action and perceptuo-motor interaction (Piaget 1967).

prompts, with only 3.7 percent of the participants correcting themselves automatically (Grandhi et al. 2011). However, these results do not clarify the relationship among more abstract representations, even those involving gestures. In an unpublished fMRI study where our volunteers watched videos of an actor speaking and gesturing, we found greater activity in brain networks for processing metaphorical and grooming gestures that have been implicated in studies on the body schema. We also found support for a generic network for amodal metaphor processing: Multimodal and monomodal metaphoric gestures do not seem to require distinct cognitive processes, but both may recruit a common semantic processing network (see also Straube et al. 2011). Although this is not definitive support, it is consistent with our proposal (to be detailed in sections 4 and 5) that gestural information creates additional features for semantic frames, reinforcing frames through manual actions.

Making a stronger claim, Barsalou has suggested that symbols are grounded in units of perception and are stored in such schematic building blocks, or perceptual symbolic units, which can be pieced into holistic frames (Barsalou 1999; Barsalou et al. 2003). In his solution of the symbol grounding problem, “perceptual symbolic units” are reminiscent of prototypical events (section 3.1) and “holistic frames” are akin to semantic frames. A criticism of Barsalou’s account is that even if symbols were grounded in modality-specific experiences, it does not automatically follow that the maintained representations have not evolved away from them. However, numerous studies have suggested that body schema and gesture influence mental representations (e.g. Cook et al. 2013).

4 Gestures Evoke Embodied Frames and Scenes: Metonymic Bases of Metaphoricity

We develop our notion of “embodied physical action frames” based on Fillmore’s (1982) idea of frames as conceptual systems originating from patterns of experience and practice. Frames have been shown to underpin not only lexical semantics but also mental spaces, metaphor, metonymy, constructions, and the understanding of discourse, to mention only a few central areas of linguistic structure and meaning construction (e.g. Coulson 2001; Dancygier & Sweetser 2014; Sullivan 2013; Ziem 2014). They represent the interlocutor’s dynamic holistic representations of the world and impose a filter in her interactions with the world and other interlocutors. We will back this view with support from neuroscience below (section 4.2). Frames do not only underpin our pragmatic action-based account of entrenched metonymic and metaphoric processes but also (i) illuminate the relation between meaningful gestural acts and the semantics of the concurrent speech and (ii) allow us

to describe how basic physical actions may be used and understood metaphorically without necessarily assuming conceptual mappings.

4.1 *Frames and Scenes Evoked in Multimodal Discourse*

Words evoke frames. Besides determining the meaning of linguistic expressions, frames also guide their use:

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: 119)

Similar to the frame functions outlined by Fillmore, gestures may provide a

“description of linguistic meanings” and are “structured in accordance with some motivating context” such as “some pattern of practices.” Gestures, too,

are often shaped by “some history of social institutions” (*Ibid.*). For instance, the TEA-DRINKING frame entails culturally defined constellations of persons, objects, and ingredients, as well as sequences of actions and interactions, which no doubt influence how one can gesturally ask someone to refill one’s tea cup.

So while frames are abstract conceptual structures, gestural frame evocation may illustrate that at least a subsection of them is grounded in physical actions. Crucially, like frames, gestures are also shaped by contexts and habits, or “scenes.” Fillmore (1977) applies this term to real-world experiences, situative factors, and non-linguistic conceptual dimensions feeding into semantic frames (see also Ziem 2014). Scenes are particularly relevant to how interlocutors construct and interpret processes of online meaning constitution: “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: 126).

Semantic frames can become salient or dominant when reinforced through repeated activation. Similarly, features of a frame can become salient through repeated activation. Processes of lexicalization (Fillmore 1982) exemplify these dynamic dimensions of frames, for example, when a neologism enters the dictionary or a novel metaphor is conventionalized. Conventional metaphoric mappings can be so salient that they dominate, even when they are irrelevant in a particular context and interfere with the appropriate semantic understanding (Glucksberg et al. 1982). Likewise, this should apply to emblems or other conventionalized or habitual gestures. This dynamic and experience-rooted nature of representations occurs even on the neuronal level: When a familiar input (frame) is disused, its mental representation loses

functionality and hence salience (Wang et al. 1995). That is, experience induces long-term structural reorganization of mental representations.

We propose that verbal gestures that recruit frame structures have a natural, pragmatically entrenched tendency to single out elements and dimensions of “prototypical instances of scenes” (Fillmore 1977: 63), in addition to primary scenes and subscenes. Gestures may metonymically profile salient dimensions of an action or object; that is, they may enact aspects of the motivating contexts of frames (see also Mittelberg 2017).

4.2 *Frames and Metonymy*

We consider frames as networks of metonymic relations that may be partially or wholly activated during frame-based reasoning, language use, and discourse understanding (e.g. Dancygier & Sweetser 2014; Fillmore 1982). An example of semantic frame activation is evoking the EMAIL frame through a simple gestural action (i.e. typing) that allows the addressee’s embodied mind to associate not only a keyboard with the movement but also possible contents of the email message and the receiver’s reply. This email example is a multimodal instantiation of what Dancygier and Sweetser (2014: 134) call “frame metonymy” and define as “the use of some entity A to stand for another entity B with which A is *correlated*” (*ibid.*, italics in the original). Drawing on Langacker’s (1987) notion of *active zone* as the profiled part of a whole, Dancygier and Sweetser (2014: 108) emphasize that a certain degree of salience is needed to clearly associate a term with a frame: “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” (e.g. “field hands”; ‘people who do mainly physical work’). This naturally applies to the gesturer’s hands, the body parts involved in both physical actions and communicative gestures and hence in evoking frames. In such cases, the following metonymic processes may apply: SALIENT PART OF FORM FOR WHOLE FORM (Barcelona 2009: 395) or SALIENT ASPECT/PHASE OF ACTION FOR WHOLE ACTION. Through frame metonymy, contiguity relationships between gesturing hands and the adjacent semanticized spaces and imaginary entities implied in their actions may become meaningful (Figures 7.1, 7.3). That is, they draw on metonymic correlations such as ACTION FOR OBJECT INVOLVED IN ACTION, PRESENTATION FOR PRESENTED, OR LOCATION FOR OBJECT (Mittelberg & Waugh 2009, 2014; Panther & Thornburg 2003).

Metonymy helps evoke relevant semantic frames (schemas) or parts of them when input is noisy or incomplete, as is common in face-to-face communication. To make sense of an environment, bottom-up sensory feature input is integrated and matched to features of semantic frames. Input might be

grouped into categories, such as discourse events, or remain as perceptual units (Barsalou 1999). Features become relevant when they can be set in a semantic frame that has been determined as relevant for a given context.

This integration or frame setting is important even at a neuronal level, where bottom-up stimulation from the environment alone is insufficient to change cortical representations. Instead, cortical changes (mental representations) depend on the interplay between bottom-up experiences and attention-modulating top-down goals (Ahissar et al. 1992).

Activating a culturally (including linguistically) defined frame includes the “co-activation” of its features. Co-activation also has bases in neuroscience where synchrony of neuronal firing creates mental representations and learning (i.e. reorganization and creation of representations). Groups of cells that ‘fire together’ have been argued to “wire together” in Hebbian’s theory of learning, providing an explanation of how contiguity can lead to mental representations. For example, separate mental representations of previously separate sensory inputs will merge when the inputs subsequently occur only in temporal synchrony (Wang et al. 1995). Similarly, the subscenes constituting a primary scene are associated in this way (see section 4.3 for further discussion of primary metaphors).

4.3 *Interaction of Metonymy and Metaphor*

Metaphor and metonymy interact to varying degrees in language and other sign systems (e.g. Jakobson 1956; Kövecze 2013; Radden 2000; see also *this volume*: Littlemore; Ruiz de Mendoza). Investigating how indexical and iconic principles jointly guide the interpretation of predominantly metaphoric gestures, Mittelberg and Waugh (2009) suggest two distinct but intertwined mappings in which metonymy leads into metaphor. For instance, to reconstruct the meaning of the gesture evoking an infinitive (Figure 7.1a), we can first assume a process of metonymic inferencing. The metonymic source – i.e. the two hands involved in the source action of holding something – points to the immediately contiguous metonymic target: the virtual object involved in the action. Second, the same imaginary object serves as the source of the metaphoric mapping whose target is the abstract notion of “infinitive” referred to verbally (see also Taub 2001: 94ff.).

Though cognitive, experiential, or functional “domains” (e.g. Croft & Cruse 2004) can also be understood in frame-semantic terms, Dancygier and Sweetser (2014) convincingly argue that, compared to domains, the structural organization of frames allows for a more systematic and fine-grained account of correlations, not only *within* a frame – giving rise to frame metonymy – but also *between* two frames that are partially mapped onto each other – giving rise to metaphor. It is central to the present perspective on frame-based multimodal

(primary) metaphor involving gesture that “[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: 137).

Also reinforcing our embodied account of frames and gestural frame evocation are brain imaging findings that have implicated the temporo-parietal junction (TPJ) not only in controlling semantic frames (Renning et al. 2013) but also in supporting the body schema. In our own unpublished neuroimaging studies, we have also found the TPJ to be involved in more complex metonymic resolution in verbal gestures. Moreover, we have found the TPJ to exert top-down modulation of semantic network areas previously implicated in the metaphoric processing of multimodal discourse involving speech and gestures, regardless of the modality in which the metaphor is expressed. Other studies have implicated the TPJ in roles that are consistent with the function of metonymy in multimodal communication, namely, in attentional control or updating top-down control (Geng & Vossel 2013). This is not surprising, as the TPJ has been argued to be a hub that integrates inputs from different sensory modalities. Thus, body schema, sensory integration, top-down attention modulation, and more complex metonymy resolution are all linked to one neuronal structure. However, as such reverse inference can only attest to the *plausibility* of our proposals, future studies focusing on the role of metonymy and frames in multimodal communication are needed. Studies could capitalize on gestures being an “exbodyment” (Mittelberg 2013b) and physical realization of meaning construction.

5 Gestural Frame Evocation: *Source Actions* as Sources of Meaning and Understanding

Throughout the chapter, we have highlighted how aspects of physical action feed into basic embodied construal operations and dynamic semantic structure. In line with the stance that interlocutors “do metaphor” while speaking, we advocate shifting the focus from object-oriented schemas, source domains, and mappings to *action-based source domains*, or what we call “source actions.” Tapping into both literal and metaphoric meanings via the action-based nature of communicative gestures, we distinguish between two types of frame structures with varying degrees of abstraction and complexity: “basic physical action and object frames” and “more complex abstract frame structures.” For the present exposition, basic action and object frames are particularly relevant.³

Basic physical action and object frames are strongly grounded in bodily experience, involving physical action and interaction with the social world and

material culture. These basic embodied frames encompass Slobin’s (1985) notion of prototypical events, Andrén’s (2010) action gestalts, Zlatev’s (2005) mimetic schemas, and what Lakoff (1987: 270–271) calls “basic-level actions” (e.g. running, walking, eating). In multimodal discourse, gesturing hands often pretend to hold, place, move, group, or else create reified abstract entities (e.g. Mittelberg 2013b, Müller 1998; Streeck 2009). Speakers evoke basic frames through such gestures and postures, through imitating walking up a steep slope with the torso leaning forward and supporting arm movements, or toasting to someone (pretending to hold an imaginary glass). Many of these actions may also function as source actions in more specific metaphoric linguistic expressions, as in “I’ll walk you through the text” (THOUGHT IS MOVEMENT) or “she grasps new concepts without problems” (UNDERSTANDING IS GRASPING; Lakoff & Johnson 1980). In each instance, two different frames that are correlated partially map onto each other (Dancygier & Sweetser 2014).

In a similar vein, Hostetter and Alibali’s (2008: 502) *gestures-as-simulated-action* framework “asserts that gestures emerge from the perceptual and motor simulations that underlie embodied language and mental imagery.” That is, schematic or abstract scenes are assumed to function as blueprints for gestures that simulate actions involving the whole body or parts of the body (see also Mittelberg 2017). When gestures simulate actions or scenes in schematic ways, the question is under what circumstances what kinds of actions gain metaphoric meaning.

5.1 Source Actions with Different Communicative and Cognitive-Pragmatic Functions

We call deeply embodied patterns of experience with the physical and social world that activate a basic physical action frame “source actions.” They range from intrinsic sensorimotor routines to actions involving people, objects, surfaces, and/or tools. As such, they may illustrate or partake in intransitive, transitive, or ditransitive clause constructions (Goldberg 1995; Hopper & Thompson 1980). In ongoing discourses, speakers may employ source actions to perform different kinds of communicative, cognitive-pragmatic tasks:

- (i) ironically enact aspects of *basic action frames* also evoked through speech (see section 5.2);
- (ii) perform a *gestural framing action* by manipulating entities evoked through speech (see section 5.3);
- (iii) carry out a *discourse-structuring activity* at a meta-level, such as sorting speech contents by placing them in different regions of gesture space.

Source actions provide cues to interlocutors’ intentions and to which aspects or moments are made salient in a dynamically evolving and intersubjectively shared multimodal discourse context (e.g. Goodwin 2011b; Jakobson 1956;



5.2 Basic Physical Action Frames Evoked through Speech and Iconic Gestures

The most essential kind of source actions encompasses basic intransitive and (di-)transitive *physical action frames* evoked through the concurrent speech and ironically simulated via gesture. Let us first consider an intransitive action (Figure 7.2).

The speaker in Figure 7.2 describes how she experienced watching a short film in which sound effects take center stage: ‘‘No, well . . . as a matter of fact I felt pulled directly into the film.’’⁴⁴ She expresses the feeling of being pulled into an emotional space, created by watching the film, with the intransitive clause, ‘‘I felt pulled directly into the film’’ while gesturing inward with both hands, palms facing her body, and pointing slightly upward and toward her torso. Interestingly, she conceptualizes the space close to her own body as the area into which she felt that she was being pulled, adopting an allocentric view. Her subtle inward movement is metaphorically contextualized through the

Mittelberg & Waugh 2014; Müller & Tag 2010; Streeck 2009). Importantly, source actions may be used and understood literally or metaphorically, depending on the interlocutors' mental frame, the subject matter, the speech content, and the overall context. Different kinds of metaphoric source actions will be illustrated next.

a. Reaching into a constituent

Figure 7.3 Explaining constituent structure

b. Pulling out a sub-constituent

discourse context: She is talking about her subjective viewing experience and emotional reaction. It may count as a metonymic ‘‘emotive’’ (Jakobson 1960) gestural enactment of a prototypical event or mimetic schema PULL, subtly instantiating the FORCE schema (‘being pulled’) as well as the image schemas IN-OUT and NEAR-FAR (Lakoff & Johnson 1999). The primary conceptual metaphor MENTAL/EMOTIONAL STATES ARE CONTAINERS may be pragmatically inferred. These interacting embodied construal mechanisms are triggered by evoking the basic action frame PULL (IN), which functions here as a source action that metonymically and metaphorically correlates physical movement and space with mental/emotional experiences and states.

The next example contains both an intransitive and a transitive construction (Figure 7.3). A linguist is explaining the hierarchical relationship between two kinds of constituents: ‘‘So, ‘invite John’s sister’ is a constituent, but we can reach into it and pull out a sub-constituent.’’ While the main constituent is metaphorically construed as a CONTAINER with a profiled interior and the sub-constituent is construed as a smaller OBJECT (to be pragmatically inferred), it is the gestural actions interacting with these imaginary constructs that are perceptually salient. These dynamic enactments function here as source actions anchoring two object-based primary metaphors via metonymic correlations between gesturing hands and contiguous spaces or objects. The speaker verbally and gesturally evokes (i) an intransitive basic action frame (REACH INTO) and the correlated (implied) spatial goal or target space (a CONTAINER), as well as (ii) a transitive basic action frame (PULL OUT) and the correlated (implied) contiguous, hand-held imaginary OBJECT (the patient). Note that the physical (iconic) source actions of reaching and pulling get a metaphorical reading through



Figure 7.2 Feeling pulled into a film

cross-modal interaction of gestural action and linguistic information about the implied imaginary, abstract categories. The ways in which semantic roles (Fillmore & Baker 2009) factor into these bimodal instances of frame evocation still requires further investigation.

5.3 Gestural Framing Actions Not Mentioned in Speech

Unlike source action frames, *gestural framing actions* typically are not specified in speech, that is, they do not enact a linguistically evoked frame. However, such iconic bodily actions can be constitutive and perceptually salient components of multimodal performances.

Returning to our first example (Figure 7.1), we see an instance of a monomodal (i.e. only bodily expressed) metaphor. Although the accompanying speech is not figurative, the gestures portray an ‘infinitive’ as a generic physical object that can be touched and manipulated through drawing on the object-oriented primary metaphors IDEAS ARE OBJECTS and CATEGORIES ARE CONTAINERS (Lakoff & Johnson 1999). Crucially, this portrayal of an abstract category is semiotically contextualized and *framed* by an enactment of holding a tangible item. In our view, such strongly grounded, prototypical transitive source actions may function as gestural framing actions that evoke the metonymically correlated, continguously handheld – or otherwise manipulated – imaginary items mentioned in the affiliate speech (Mittelberg 2006; Mittelberg & Waugh 2009).

One needs to remember that, even with the aid of body schema or body image, no single gesture has an unequivocal interpretation. Without the affiliate speech, interpreting gestures is rendered even more difficult: An ambiguous bodily sign does not effectively evoke the appropriate semantic frames that lead to understanding the intended meaning. Moreover, ineffective or weak, and hence non-salient, metaphors tend to be ignored (Glucksberg et al. 1982). Thus, in order for monomodal metaphoric gestures to be communicatively effective, they would need to recruit strongly embodied primary metaphors (and the related image and force schemas), for instance, via metonymically correlated gestural framing actions.

5.4 Some Implications

The fact that non-emblematic, spontaneous coverbal gestures seldom have unique interpretations reminds us of the importance of how the overall sum (i.e. the intended meaning communicated) in multimodal communication is greater than the sum of its parts (i.e. the meanings derived from each modality, taken separately). This phenomenon is known as “superadditivity” in physiological and neuroscience research of multisensory integration. Superadditivity

also hints at the workings of metonymy or frames: “Parts” or features can map onto more complete “parts” (themselves semantic frames, combining a number of features), which, in turn, are parts of yet more complex, holistic semantic frames. Hence, it might appear that the sum of the parts leads to greater activation than the parts in isolation. In multimodal discourse, understanding is a distributed (inter-)subjective, cognitive process requiring grounding in experientially motivated frames and scenes that are strongly perception- and action-based. Monomodal metaphoric gestures seem to rely on this process even more heavily.

Along with the semantic frames involved in processing spoken utterances, some additional frames seem to be needed to process actions in multimodal discourse. These frames, like gestures, could be related hierarchically. As features are weighted according to how essential or salient they are for an entrenched pattern of experience, this hierarchy seems to be determined by the capacity of a given feature or scene to cue for a particular frame (see also Dancyger & Sweeter 2014). This lets us assume that the more (salient) features are profiled, the more complete and hence more strongly the frame is activated. The intended meaning of coverbal gestures is strongly determined by the context, which is often greatly shaped by the accompanying speech (see Müller, *this volume*). However, when the speech does not aid in resolving the intended meaning of a gesture, its decodability relies on its ability to convey features that are salient or embodied enough to be able to activate frames on their own. Without attentional control and a mechanism as a conduit to handle noisy input, these features remain as noise. Metonymy provides a means to “fuzzy match” in order to activate relevant frames with less effort on the part of the speaker. Without cues in the affiliate speech, people seem to follow a set of strategies, based on body schema and action routines that are part of these schemata, to interpret gestures representing, for instance, literal transitive actions (e.g. Grandhi et al. 2012).

We have argued that metaphoric understandings expressed through gestures, especially gestural framing actions involving metaphorically construed items, tend to – via metonymy – involve primary metaphors (in particular for monomodal metaphoric gestures) or other very salient, strongly embodied conceptual metaphors that recruit primary scenes (Grady & Johnson 2002), prototypical events (Slobin 1985), basic-level actions (Lakoff 1987), basic image and force schemas, certain mimetic schemas (Zlatev 2005), and in particular, basic embodied frames and scenes (Fillmore 1977).

6 Concluding Remarks

We have laid out the foundations of a frame-based account of gestural action. The *basic action frames* and *source actions* we have presented may function

as bodily sources of meaning, experience, and understanding (Johnson 2007) in multimodal discourse where speech, gestures, and other modalities contextualize one another. As emphasized, metonymic modes allow us to infer full action routines, emotive qualities, and interactive dimensions from schematic gestural articulations, as well as correlated imaginary objects or traces from visible gestural actions (Mittelberg & Waugh 2009, 2014).

When such gestural actions occur in discourses about abstract knowledge domains, mental processes, affective states, or other kinds of subjective experiences, they may receive a *contextually driven* metaphorical interpretation (see also Müller & Cienki 2009). This resonates with research on language processing showing that certain linguistic contexts cause a metaphorical reading of verbal expressions that, in principle, could have either a literal or a figurative meaning (e.g. Giora 2008). Once interlocutors are in an abstract or emotional mind-set, gestural actions, including framing actions, can mediate between physical and cognitive processes and thereby provide experiential grounding, for instance, through “imaginary material anchors” (Hutchings 2005), such as imagined objects, even if the concurrent speech is non-figurative.

Our frame-based account of how gestural source actions may profile “prototypical instances of scenes” (Fillmore 1977: 63) – or enact relevant aspects of “scenes basic to human experience” (Goldberg 1995: 5) – allows us to better understand why metaphoricity that is monomodally expressed through the communicating body relies on metonymy and tends to involve primary metaphors (Kövecses 2013). Working out how scenes and frames may also motivate multimodal constructions is one ensuing step of inquiry (e.g. Barcelona 2009; Mittelberg 2017; Sullivan 2013).

The gestural and neurocognitive patterns discussed in this chapter further attest to the psychological and semiotic reality of internalized conceptual operations, entrenched metaphorical understandings, and social practices (Gibbs 2006a; *this volume*: Gibbs; Winter & Matlock; Müller). Complex conceptual networks do not necessarily translate into neural network fingerprints. Although grounded strongly in bodily experience, the interplay of this bodily experience with other contextual factors makes linguistic and multimodal cognition a multidimensional process requiring further studies before an integrative understanding can be gained. Here we have outlined a theoretical account, anchored in empirical findings across different disciplines, in order to help shape goals of future cognitive-linguistic and neuroscience research into linguistic cognition, discourse processes, and the role of the body.

Perhaps the most intriguing question re-arising from these body-centered and action-based insights is whether, or rather in which cases, we actually need to assume conceptual metaphorical mappings when deeply embodied, basic frame-based correlation metaphors are recruited for communicative purposes via their metonymic bases. From a discourse pragmatics perspective, it indeed

NOTES AND ACKNOWLEDGMENTS

* We thank the editor and an anonymous reviewer for a very helpful discussion of a number of critical points. The research presented in this chapter was supported by the Excellence Initiative of the German Federal and State Governments.

1 This approach builds on Fillmore’s notion of “frame,” as developed before the FrameNet project (Fillmore & Baker 2009).

2 German original: “Sie schauen einen (...) Infinitiv an, können sie sagen *ah*, das ist, (...) Präsens, Perfekt, Plusquamperfekt?”

3 For a discussion on gestures evoking complex, abstract frame structures, such as discourses and theories, see Mittelberg (2017).

4 German original: “Nö, das also ... also ich fühlte mich schon direkt reingezogen in den Film.”