



Social Interaction. Video-Based Studies of Human Sociality.
2022 Vol. 5, Issue 1
ISBN: 2446-3620
DOI: 10.7146/si.v5i2.130875

Social Interaction

Video-Based Studies of Human Sociality

On Gestalts and their Analytical Corollaries: A Commentary to the Special Issue

Elwys De Stefani

University of Heidelberg & KU Leuven

1. Introduction

This article is a comment on the contributions to the current Special Issue in the light of the guest editors' declared aim to gain insight into "how bodily practices feature in action formation and action ascription in multilingual interaction" (Pirainen-Marsh, Lijla & Eskildsen, 2022/this issue). All authors rely on conversation analytic research standards, and they all use video data of interactions between L1 and L2 speakers of a variety of languages. The contributions hence offer a welcome investigation into practices deployed by L1 and L2 speakers in settings that are different from classroom interaction, on which there is a plethora of research from diverse vantage points from the 1960s onwards (e.g., Barnes, 1969; Sinclair & Coulthard, 1975), and which, with respect to second-language acquisition (SLA), has been studied from a conversation analytic perspective under the label *CA for SLA* (Markee & Kasper, 2004) or *CA-SLA* for almost twenty years now (for a recent overview, see Kunitz, Markee & Sert, 2021). Conversation analytic approaches to classroom interaction — both

in L1 and L2 settings — are still flourishing and have increasingly taken embodied behavior into analytical consideration (see Majlesi, 2015; Hall & Looney, 2019; Eskildsen, 2021, among many others). Interest in second-language learning in the wild has been growing (Hellermann et al., 2019), and more and more attention is being devoted to multimodal interactions between L1 and L2 speakers in ordinary settings of interaction from the perspective of researchers studying L2 acquisition (Svennevig, 2018; Greer, 2019; Lilja & Piirainen-Marsh, 2019; Greer & Wagner, 2021; Kurhila, Kotilainen & Lehtimaja, 2021). In general, research on interaction between individuals with unequal language competencies is not seldom and addresses multilingual professional settings (e.g., Mondada, 2012a; Hazel & Svennevig, 2018), ordinary encounters between unacquainted individuals (e.g., Torras & Gafaranga, 2002), but also interactions between aphasic and non-aphasic persons (e.g., Goodwin, 1995; Wilkinson, 2013).

The authors of this Special Issue examine a variety of settings where L2 speakers interact with L1 speakers of the language of their encounter, except for one study involving only L2 speakers (Skogmyr Marian & Pekarek Doehler, 2022/this issue). The settings cover conversation cafés for language learners with (Majlesi, 2022/this issue; Kunitz & Majlesi, 2022/this issue) or without (Skogmyr Marian & Pekarek Doehler, 2022/this issue) L1 speakers, dinner table interactions in families hosting foreign exchange students (Greer, 2022/this issue; Lilja & Eskildsen, 2022/this issue), interaction with an au-pair (Frick & Palola, 2022/this issue), a cooking class for migrants (Lilja & Piirainen-Marsh, 2022/this issue), and video-based online interactions between L1 and L2 speakers of German (Uskokovic & Taleghani-Nikazm, 2022/this issue). All of these settings enable the contributors to document how participants orient, to different degrees, to the appropriateness or correctness of the L2 speakers' language productions. However, the main focus of the Special Issue being on *bodily practices*, *action formation*, and *action ascription in multilingual interaction*, the acquisitional dimension of the observed practices is less predominantly discussed by the authors, with the exception of Skogmyr Marian & Pekarek Doehler, who show how word search practices change as learners become more proficient in the target language over time, and Uskokovic & Taleghani-Nikazm, who document word searches used by L2 speakers engaged in video calls. Some articles focus on practices deployed by proficient speakers in their interaction with learners, in particular in instructional settings, such as cooking classes (Lilja & Piirainen-Marsh), but also in the L1 speakers' displays of their higher linguistic proficiency (Majlesi; Lilja & Eskildsen; Kunitz & Majlesi).

This discussion measures the contributions of this Special Issue against the foci defined by the editors.

2. Bodily practices

While bodily practices are central to all the contributions, the nature of the phenomena described, the granularity of the analyses, and the conceptual and terminological toolkits vary considerably. Most articles focus indeed on hand gestures or related conduct, such as the raised index finger (Uskokovic & Taleghani-Nikazm), the Japanese *gassho* gesture (Greer), gesture holds (Skogmyr Marian & Pekarek Doehler), or gestural matching (Majlesi). Some articles examine gestures based on prior conceptualization, e.g., as “depictive gestures” (Skogmyr Marian & Pekarek Doehler; Lilja & Piirainen-Marsh). All articles examine bodily behavior from a holistic perspective, taking into analytical consideration mainly gesture, gaze, and talk.

The conceptual framework that authors mobilize for identifying and describing gestures suggests an effortless distinction between “pragmatic gestures” (Streeck, 2009; Kendon, 2004) and “depictive gestures” (Streeck, 2009) or “iconic gestures” (Ekman & Friesen, 1969), which recalls McNeill’s (1985) distinction between “referential” and “discourse-oriented” gestures. This distinction is highly relevant to Skogmyr Marian & Pekarek Doehler’s contribution, which finds that word searches with co-occurring depictive gestures invite co-participants to cooperate in the word search, whereas speakers who produce pragmatic gestures (or self-touch) while engaged in a word search display that they are holding the floor and do not invite help from other co-participants. The different treatment of depictive gestures vs. non-depictive manual action is indeed backed by prior research, e.g., by Goodwin (1986), who showed that recipients tend to gaze to depictive gestures but not to self-touch practices.

However, the categorization of manual action as a (non-)depictive gesture is sometimes arguable. While in many cases the depictive or iconic use of a gesture is apparent, in others it is open to question. For instance, in Skogmyr Marian & Pekarek Doehler’s Excerpt 5, the participant Aurelia performs a hand movement while engaged in a word search (line 3), and the authors describe it as “what can be seen as a depictive gesture.” The hesitant formulation is warranted since the movement of the hands could also “be seen” as a pragmatic gesture. It would have been highly interesting to understand on what grounds the authors were able to identify this particular gesture as an instance of a depictive gesture, as it is an analytical problem many scholars working on gesture have to face. On the one hand, gestures are never “purely” pragmatic, depictive, deictic, etc. (see Goodwin, 2003), on the other hand, especially pragmatic gestures have been described as an “unruly bunch: speakers show all manner of idiosyncrasies in making them” (Streeck, 2009: 181). And authors sometimes tend to assign meaning to them, such as Frick and Palola, who describe arm-spreading as a gesture “associated with resignation, helplessness, and the meaning ‘I don’t know’” (Excerpt 2, line 3). If it is true that some gestures are culture-specific (Greer; see also De Jorio, 1832), then it would be necessary to know in which

culture (or language) arm-spreading is “associated” with the meaning described by the authors.

Rather than focusing on a gestural function (such as depictive or pragmatic), some authors identify specific shapes of hand movements (such as the *gassho* gesture described by Greer, or the raised index finger analyzed by Uskokovic & Taleghani-Nikazm), which dispenses them from having to refer to a functional gesture typology, thereby making the identification of the target gesture easier. However, since gestures are always indexical, their understanding is sensitive to their situated use, to how they temporally and sequentially relate to enviroing talk. Hence, while Uskokovic & Taleghani-Nikazm very thoughtfully examine how L2 speakers of German engaged in video-mediated interaction may raise their index in word searches, they show occurrences in which participants might do something more (or something else) than “just” searching for a word. The authors orient to that, for instance, when they write that in Excerpt 3, the problem for the L2 speaker is not a missing word, but the conversion from the US system of measurements to the metric system, or with regard to Excerpt 5, which is described as an “incomplete word-search.”

The above observations hint towards two kinds of difficulties faced by many interactionally oriented researchers working with naturalistic data when they examine bodily behavior. First, while the transcripts often show a precise indication of how gesture, gaze and other behavior temporally relate to talk, the analyses do not always show the same level of granularity. Second, while the next-turn proof procedure (Sacks, Schegloff & Jefferson, 1974: 729) has been described as “the most basic tool used in CA to ensure that analyses explicate the orderly properties of talk as oriented-to accomplishments of participants, rather than being based merely on the assumptions of the analyst” (Hutchby & Woofitt, 1998: 15), it is less readily accessible for analysts examining manual behavior. Therefore, researchers often (have to) resort to more or less plausible assumptions about the bodily behavior they observe. Gestures “seem to index,” “can be seen as,” “may be designed to” (Skogmyr Marian & Pekarek Doehler), “denote” (Greer), are “associated with” (Frick & Palola), “may be indicating” (Uskokovic & Taleghani-Nikazm) or “signal” (Lijla & Eskildsen). These often modalized descriptions show the authors’ need to assign one specific meaning or function to a gesture — a way of thinking, which, paradoxically, is in stark contrast to the idea that gesture is accountable and thus interpretable only in the temporal, sequential and situated environment of its occurrence.

Virtually all the authors of the current Special Issue subscribe to the idea that gestures are deployed within “multimodal action packages” or “multimodal Gestalts” (Mondada, 2012b). As Stukenbrock (2021) has shown, the idea that vocal and visual behavior have to be analyzed as forming a Gestalt or package (both terms are found already in Heath, 1986) has been swiftly embraced by interactionally oriented researchers working with video data. However, what researchers treat as multimodal Gestalts and how such Gestalts are identified

varies considerably, especially with respect to the granularity of their description. According to Mondada (2014: 140) *complex multimodal Gestalts* are “both specifically adjusted and systematically ordered,” they are “packaged in an emergent, incremental, dynamic way.” They are, of course, not just recurrent, uniform reproductions of the “same” multimodal arrangement. They are sensitive to the local and sequential environment at hand, and they are highly adaptable — with regard to their temporal deployment, the coordination of the different modalities, and their manifestation in space. And they are recognizable, for interactants, as such. The adaptability of Gestalts, the idea that Gestalts are not just the sum of their components, but that the way in which the different components are arranged make *that* Gestalt perceivable is a central thought of the founders of Gestalt theory (Wertheimer, 1985 [1924]). From this point of view, Uskokovic & Taleghani-Nikazm’s study neatly shows how a raised index finger produced in a specific sequential environment (self-initiated repair) is perceived as participating in a word search (and not, e.g., as a pointing gesture). Similarly, Greer’s contribution shows a recurrent systematic arrangement of a manual, a bodily and a vocal practice, which participants locally assemble as a display of an apology.

3. Action formation and ascription

The main claim of the authors is that social actions are multimodally accomplished. The apparent obviousness of this assertion resonates with Schegloff’s (2007: xiv) well-known description of the action formation problem: “[...] how are the resources of the language, the body, the environment of the interaction, and position *in* the interaction fashioned into conformations designed to be, and to be recognized by recipients as, particular actions [...]?” Schegloff exemplified what he regarded as “particular actions” with a list of action types that have been thoroughly studied (not only) by conversation analysts (requesting, complaining, agreeing, etc.). What counts as an action, how actions are displayed and how individuals ascribe (Levinson, 2013) an action to an observable behavior is, however, an ongoing debate (e.g., Enfield & Sidnell, 2017), which, it seems, relies heavily on conceptualizations that assign vocal language precedence over embodied conduct (with the notable exception of ethnomethodological studies on signed and tactile languages, e.g., McIlvenny, 1995, Iwasaki et al., 2019 and studies such as Heath & Luff, 2021, which focus mainly on gesture and object manipulation from an interactional perspective). Notoriously, when Sacks started analyzing recorded conversations, “it wasn’t from any large interest in language” (Sacks, 1992, 1: 622). Yet he developed fabulous tools enabling us to understand how social action is achieved and discerned through talk, while emphasizing at the same time the centrality of visual perception in human interaction: “For Members, activities are observables. They see activities” (Sacks, 1992, 1: 119). The fact that Sacks spoke of “activities” (rather than “actions”) is noteworthy. “Activities” has the advantage of suggesting a more or less complex conduct,

whereas the notion of “action” seems to imply that participants are able to identify what might be called ‘praxeological units’ — and indeed, the recent conversation analytic studies on requests and other first actions are in line with this way of thinking. At the same time, however, numerous researchers in ethnomethodology and conversation analysis agree that human behavior cannot always be parsed into single, neatly distinguishable actions. Hence the need to describe “double-barreled” (Schegloff, 2007: 76) or “composite” (Rossi, 2018) actions and Levinson’s (2013) suggestion that participants ascribe a “primary” or “major” action to observable conduct (see, e.g., Lilja & Eskildsen’s contribution on what they call “repairing-for-teasing”). This resonates, within linguistics, with the notion of “indirect speech acts” developed in pragmatic theory (Searle, 1975).

On the other hand, Sacks drew attention to the fact that activities are “observed” and “seen.” Perception is, of course, at the core of the Gestalt-theoretic interest (Koffka, 1922), as well as of phenomenological positions in general (Merleau-Ponty, 1945; Wittgenstein, 1953; Gurwitsch, 1957). By grounding their analyses on recorded audio and video data, interactionally oriented researchers were able to explore the temporal coordination and sequential organization of observable human activities (Goodwin, 1979; Heath, 1986). Temporality is of course fundamental to the formation and recognition of “multimodal Gestalts,” as the contributors to the current Special Issue also affirm. Actions not only unfold in time, but are also recognized in time, even before they have been brought to conclusion. Actions are recognized early (see Deppermann, Mondada & Pekarek Doehler, 2021). According to Levinson (2013: 103), “[t]he challenge for participants, then, is to assign at least one major action to *a turn they have only heard part of so far*” (emphasis added). The task for participants in co-present interaction is probably even more complex: Not only do different modalities have different temporalities, but individuals may be engaged in diverse courses of actions with the various resources physically available to them. Hairdressers may talk to their clients while taking care of their hair, driving instructors may chat with trainee drivers while monitoring their traffic conduct (De Stefani & Horlacher, 2018). Given the principal possibility of multiactivity, the varying “temporal orders” (Mondada, 2014) with which participants organize their courses of actions, and the fact that they may prioritize one course of action over another (in particular when these are carried out with the same physical resources, such as eating and talking), assigning “at least one major action” to the embodied conduct they witness is a highly complex task for interactants. How it is possible that we “see” that different modalities coalesce into a finely organized “multimodal Gestalt” calls for further investigation. A fine-grained analysis of the different temporalities, of whether they relate to each other in terms of simultaneity or successivity might be one analytical starting point.

4. Multilingual interaction

Multilingual interaction is approached in a broad sense by the contributors to this Special Issue. It covers not only interactions that are carried out in different languages, but also interactions between participants with unequal linguistic backgrounds, but actually conversing in one language only (see also Piirainen-Marsh, Lilja & Eskildsen, 2022/this issue). In fact, the only contribution offering a multilingual setting of interaction is the one authored by Frick & Palola, which, however, does not focus on the participants' language choices.

By analyzing self-repair practices of L2 speakers, the papers by Skogmyr Marian & Pekarek Doehler and Uskokovic & Taleghami Nikazm show that these orient to a “one language only” policy. Language policing having been observed mainly in classroom interaction (Amir & Musk, 2013), the fact that the speakers observed in the above-mentioned contributions orient to the “one language only” principle shows that they are treating the interactions in which they are engaged as constituting a learning environment. Contrary (perhaps) to the L2-classroom, in this setting, the orientation towards normative language use is less pervasive: word searches seem to aim at respecting the “one language only” policy rather than at producing grammatically and normatively “correct” talk (see, e.g., Skogmyr Marian & Pekarek Doehler's Ex. 1, where the speaker engaging in a word search eventually says “il y a beaucoup de [] pour faire shopping” / “there are many [] to do shopping”, without ever articulating the searched-for noun and without producing the idiomatically expected partitive “pour faire *du* shopping”). This does not mean, of course, that normative other-corrections may not be present in the corpus analyzed. It is striking, however, that in word searches designed by participants to be solved on their own, grammatical normativity does not seem to be at stake. The question then is, what exactly are speakers learning? In her article on word searches between L1 and L2 speakers, Brouwer (2003: 542) described occasions in which “the other participant is invited to participate in the word search” as “language learning opportunities.” Whether and in what way the practices observed by Skogmyr Marian & Pekarek Doehler and Uskokovic & Taleghami Nikazm are such opportunities might be further explored. The former's longitudinal perspective is certainly an excellent approach to answer this question (see also Eskildsen, 2018).

Word searches initiated by L2 speakers who invite a collaborative solution are at the center of Majlesi's contribution, where (more) proficient and L1 speakers clearly display an orientation towards normativity. Both in Majlesi's self-initiated repair sequences and in Lilja & Eskildsen's other-initiated repair sequences, by which proficient speakers target what they treat as inadequate pronunciation by L2 speakers, the participants orient towards the “correct” pronunciation, lexical choice, etc., thereby treating their interactions as aiming at learning and teaching the “proper” way to speak (see also Kunitz & Majlesi).

Greer's contribution — focusing on Japanese learners of English articulating (English) tokens of apology while producing the (Japanese) *gassho* gesture — offers a particular “contact” phenomenon resulting from the combination of interactional resources available in Japanese and English for the accomplishment of an apology. While he describes its occurrence as delivered as a “complex multimodal Gestalt” (Mondada, 2014), he also shows that (non-Japanese) recipients might or might not perceive it as a Gestalt but rather as a verbal apology delivered with a gesture that is “part of that action.” In other words, English-speaking recipients do not “need” it in order to see that an apology is being accomplished. Greer hence examines the production side of Gestalts, whereas the focus of Gestalt theorists was chiefly on their perception.

5. A final word

The analysis of video data has enabled (or rather, obliged) researchers to address the complexity of social actions from a holistic perspective. These developments go hand in hand with the problem of representing, for readers and viewers, the phenomena that are at the core of the analysis. *Social Interaction. Video-Based Studies of Human Sociality* provides authors with the possibility of making available the video excerpts on which they based their analysis. However, practices of anonymization tend to make the relevant features (e.g., eye gaze) unavailable. Some authors (Lilja & Eskildsen; Lilja & Piirainen) sought to maximize the visibility of the relevant phenomena for readers by offering two transcription versions of the same excerpt, one based on Mondada's (2018) conventions, highlighting how different modalities sequentially and temporally relate to each other, and one based on Laurier's (2014) comic strip representation, which rather highlights the visibility of situational aspects. Their aim, of course, is to make accessible for readers both the temporal organization of multimodally accomplished actions and their visible features. The notion of “multimodal Gestalt” has been widely used in recent years for referring to such actions, and it features prominently also in the contributions to this Special Issue. The authors' considerations have led me to catch a glimpse of the analytical challenges the notion implies. While conversation analytic research has focused so far on how individuals assemble vocal and bodily resources to form a Gestalt-like, supposedly meaningful whole, how recipients ascribe meaning (or action, if you will) to such wholes while they emerge, is a question future research might delve into. It appears that more and more scholars embrace an understanding of social life as built on sequences of multifaceted actions, rather than (only) on sequences of discrete “praxeological units” such as question/answer, greeting/greeting etc. In this scheme, the consideration of the “subtleties of glance, of gesture, of tone”, which Wittgenstein (1953, Part IIxi, 228) identified as “imponderable evidence” is essential for the analysis of face-to-face interactions. “The question is: what does imponderable evidence *accomplish?*” (ibid., emphasis in original).

References

- Amir, A. & Musk, N. (2013). Language policing: Micro-level language policy-in-progress in the foreign language classroom. *Classroom Discourse*, 4(2), 151–167.
- Barnes, D. (1969). Language in the secondary classroom. In D. Barnes, J. Britton, and H. Rosen (Eds.), *Language, the learner and the school*. London, UK: Penguin, 11–77.
- Brouwer, C. E. (2003). Word searches in NNS-NS interaction: Opportunities for language learning? *The Modern Language Journal*, 87(4), 534–545.
- De Jorio, A. (1832). *La mimica degli antichi investigate nel gestire napoletano*. Napoli: Stamperia e cartiere del Fibreno.
- De Stefani, E. & Horlacher, A.-S. (2018). Mundane talk at work: Multiactivity in interactions between professionals and their clientele. *Discourse Studies*, 20(2), 221–245.
- Deppermann, A., Mondada, L. & Pekarek Doehler, S. (2021). *Early responses in human communication*. Special Issue of *Discourse Processes*, 58(4).
- Enfield, N. J. & Sidnell, J. (2017). *The concept of action*. Cambridge, UK: Cambridge University Press.
- Eskildsen, S. W. (2018). 'We're learning a lot of new words': Encountering new L2 vocabulary outside of class. *The Modern Language Journal*, 102(Supplement), 46–63.
- Eskildsen, S. W. (2021). Embodiment, semantics and social action: The case of object-transfer in L2 classroom interaction. *Frontiers in Communication*, <https://doi.org/10.3389/fcomm.2021.660674>
- Frick, M., & Palola, E. (2022/this issue). Deontic Autonomy in Family Interaction: Directive Actions and the Multimodal Organization of Going to the Bathroom. *Social Interaction. Video-Based Studies of Human Sociality*, 5(1). <https://doi.org/10.7146/si.v5i2.130870>
- Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. In: G. Psathas (Ed.), *Everyday language. Studies in ethnomethodology*. New York, Irvington Publishers: 97–121.
- Goodwin, C. (1986). Gesture as a resource for the organization of mutual orientation. *Semiotica*, 62(1–2), 29–49.
- Goodwin, C. (1995). Co-constructing meaning in conversations with an aphasic man. *Research on Language and Social Interaction*, 28(3), 233–260.

- Goodwin, C. (2003). Pointing as a situated practice. In: S. Kita (Ed.), *Pointing: Where language, culture and cognition meet*, Mahwah: Laurence Erlbaum, 217–241.
- Greer, T. (2019). Noticing in the wild. In J. Hellermann, S. W. Eskildsen, S. Pekarek Doehler, & A. Piirainen–Marsh (Eds.), *Conversation analytic research on learning-in-action: The complex ecology of second language interaction 'in the wild'* (pp. 131–158). Cham: Springer
- Greer, T. (2022/this issue). Multimodal action formation in second language talk: Japanese speakers' use of the Gassho gesture in English apology sequences. *Social Interaction. Video-Based Studies of Human Sociality*, 5(1). <https://doi.org/10.7146/si.v5i2.130868>
- Greer, T. & Wagner, J. (2021). The interactional ecology of homestay experiences: Locating input within participation and membership. *Second Language Research* <https://doi.org/10.1177/02676583211058831>
- Gurwitsch, A. (1957). *Théorie du champ de la conscience*. Paris: Desclée de Brouwer.
- Hall, J. K. & Looney, S. D. (Eds.) (2019): *The embodied work of teaching*. Bristol, UK: Multilingual Matters.
- Hazel, S. & Svennevig, J. (2018). Multilingual workplaces: Interactional dynamics of the contemporary international workforce. *Journal of Pragmatics*, 126, 1–9.
- Heath, C. (1986). *Body movement and speech in medical interaction*. Cambridge, UK: Cambridge University Press.
- Heath, C. & Luff, P. (2021). Embodied action, projection and institutional action: The exchange of tools and implements during surgical procedures. *Discourse Processes*, 58(3), 233–250.
- Hellermann, J., Eskildsen, S. W., Pekarek Doehler, S., & Piirainen-Marsh, A. (Eds.). (2019). *Conversation analytic research on learning-in-action: The complex ecology of second language interaction 'in the wild'*. Cham: Springer.
- Hutchby, I. & Woofitt, R. (1998). *Conversation analysis. Principles, practices and applications*. Cambridge, UK: Polity Press.
- Iwasaki, S. et al. (2019). The challenges of multimodality and multi-sensoriality: Methodological issues in analyzing tactile signed interaction. *Journal of Pragmatics*, 143, 215–227.

- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge, UK: Cambridge University Press.
- Koffka, K. (1922). Perception: An introduction to the Gestalt-Theorie. *The Psychological Bulletin*, 19(10), 531–585.
- Kunitz, S., & Majlesi, A. R. (2022/this issue). Multimodal gestalts in reformulating practices in language cafés. *Social Interaction. Video-Based Studies of Human Sociality*, 5(1). <https://doi.org/10.7146/si.v5i2.130873>
- Kunitz, S., Sert, O., & Markee, N. (Eds.) (2021). *Classroom-based conversation analytic research: Theoretical and applied perspectives on pedagogy*. Cham: Springer.
- Kurhila, S., Kotilainen, L., & Lehtimaja, I. (2021). Orienting to the language learner role in multilingual workplace meetings. *Applied Linguistics Review*. Advance online publication. <https://doi.org/10.1515/applirev-2021-0053>
- Laurier, E. (2014). The graphic transcript: Poaching comic book grammar for inscribing the visual, spatial and temporal aspects. *Geography Compass*, 8(4), 235–248.
- Levinson, S. C. (2013). Action formation and ascription. In: J. Sidnell & T. Stivers (Eds.), *The handbook of conversation analysis*. Malden, MA: Blackwell, 103–130.
- Lilja, N., & Eskildsen, S. W. (2022/this issue). The embodied work of repairing-for-teasing in everyday L2 talk. *Social Interaction. Video-Based Studies of Human Sociality*, 5(1). <https://doi.org/10.7146/si.v5i2.130872>
- Lilja, N., & Piirainen-Marsh, A. (2022/this issue). Recipient design by gestures: Depictive gestures embody actions in cooking instructions. *Social Interaction. Video-Based Studies of Human Sociality*, 5(1). <https://doi.org/10.7146/si.v5i2.130874>
- Majlesi, A. R. (2015). Matching gestures: Teachers' repetitions of students' gestures in second language classrooms. *Journal of Pragmatics*, 76, 30–45.
- Majlesi, A. R. (2022/this issue). Gestural matching and contingent teaching: Highlighting learnables in table-talk at language cafés. *Social Interaction. Video-Based Studies of Human Sociality*, 5(1). <https://doi.org/10.7146/si.v5i2.130871>
- Markee, N. & Kasper, G. (2004). Classroom talks: An introduction. *The Modern Language Journal*, 88, 491–500.

- McIlvenny, P. (1995). Seeing conversations: Analyzing sign language talk. In: P. Ten Have & G. Psathas (Eds.), *Situated order: Studies on the social organisation of talk and embodied activities*. Washington, DC: University Press of America, 129–150.
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review*, 92(3): 350–371.
- Merleau-Ponty, (1945). *Phénoménologie de la perception*. Paris: Gallimard.
- Mondada, L. (2012a). The dynamics of embodied participation and language choice in multilingual meetings. *Language in Society*, 41(2), 213–235.
- Mondada, L. (2012b). Deixis: An integrated interactional multimodal analysis. In P. Bergmann & J. Brenning (Eds.), *Interaction and usage-based grammar theories: What about prosody and visual signals?* Berlin: De Gruyter, 173–206.
- Mondada, L. (2014). The temporal orders of multiactivity: Operating and demonstrating in the surgical theatre. In: P. Haddington, T. Keisanen, L. Mondada and M. Nevile (Eds.), *Multiactivity in social interaction: Beyond multitasking*. Amsterdam/Philadelphia, PA: John Benjamins, 33–75.
- Mondada, L. (2018). Multiple temporalities of language and body in interaction: Challenges for transcribing multimodality. *Research on Language and Social Interaction*, 51(1), 85–106.
- Piirainen-Marsh, A., Lilja, N., & Eskildsen, S. W. (2022/this issue). Bodily practices in action formation and ascription in multilingual interaction: Introduction to the special issue. *Social Interaction. Video-Based Studies of Human Sociality*, 5(1). <https://doi.org/10.7146/si.v5i2.130866>
- Rossi, G. (2018). Composite social actions: The case of factual declaratives in everyday interaction. *Research on Language and Social Interaction*, 51(4), 379–397.
- Sacks, H. (1992). *Lectures on conversation*. Malden, MA: Blackwell.
- Sacks, H., Schegloff, E. A. & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735.
- Schegloff, E. A. (2007). *Sequence organization in interaction*. Cambridge, UK: Cambridge University Press.

- Searle, J. (1975). Indirect speech acts. In: P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics. Volume 3: Speech acts*. New York: Academic Press, 59–82.
- Sinclair, J. & Coulthard, M. (1975). *Towards an analysis of discourse*. Oxford, UK: Oxford University Press.
- Skogmyr Marian, K. (2021). Initiating a complaint: Change over time in French L2 speakers' practices. *Research on Language and Social Interaction*, 54(2), 163–182.
- Skogmyr Marian, K., & Pekarek Doehler, S. (2022/this issue). Multimodal word-search trajectories in L2 interaction: The use of gesture and how it changes over time. *Social Interaction. Video-Based Studies of Human Sociality*, 5(1). <https://doi.org/10.7146/si.v5i2.130867>
- Streeck, J. (2009). *Gesturecraft: The manufacture of meaning*. Amsterdam/Philadelphia: John Benjamins.
- Stukenbrock, A. (2021). Multimodal gestalts and their change over time: Is routinization also grammaticalization? *Frontiers in Communication*, <https://doi.org/10.3389/fcomm.2021.662240>
- Svennevig, J. (2018). “What’s it called in Norwegian?” Acquiring L2 vocabulary items in the workplace. *Journal of Pragmatics*, 126, 68–77.
- Torras, M.-C. & Gafaranga, J. (2002). Social identities and language alternation in non-formal institutional bilingual talk: Trilingual service encounters in Barcelona. *Language in Society*, 31(4), 527–548.
- Uskokovic, B., & Talehgani-Nikazm, C. (2022/this issue). Talk and embodied conduct in word searches in video-mediated interactions. *Social Interaction. Video-Based Studies of Human Sociality*, 5(1). <https://doi.org/10.7146/si.v5i2.130876>
- Wertheimer, M. (1985 [1924]). Über Gestalttheorie. *Gestalt Theory*, 7(2), 99–120.
- Wilkinson, R. (2013). Gestural depiction in acquired language disorders: On the form and use of iconic gestures in aphasic talk-in-interaction. *Argumentative and alternative communication*, 29(1), 68–82.
- Wittgenstein, L. (1953). *Philosophische Untersuchungen*. Oxford, UK: Basil Blackwell.