



# Multimodal Gestalts and Their Change Over Time: Is Routinization Also Grammaticalization?

Anja Stukenbrock\*

*Chair of German Linguistics, University of Heidelberg, Heidelberg, Germany*

## OPEN ACCESS

### Edited by:

Simona Pekarek Doehler,  
Université de Neuchâtel, Switzerland

### Reviewed by:

Ali-Reza Majlesi,  
Stockholm University, Sweden  
Makoto Hayashi,  
Nagoya University, Japan

### \*Correspondence:

Anja Stukenbrock  
anja.stukenbrock@gs.uni-  
heidelberg.de

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Communication

**Received:** 31 January 2021

**Accepted:** 17 September 2021

**Published:** 04 November 2021

### Citation:

Stukenbrock A (2021) Multimodal  
Gestalts and Their Change Over Time:  
Is Routinization Also  
Grammaticalization?  
*Front. Commun.* 6:662240.  
doi: 10.3389/fcomm.2021.662240

Recently, the claim was put forward that grammar emerges from embodied conduct. This has led to a discussion in multimodal conversation analysis and interactional linguistics whether the routinization of embodied actions can be described in terms of grammar and grammaticalization. While particular items such as exophoric demonstratives and gestures are routinely delivered as multimodal constructions, i.e., as part of grammar, it is debatable whether this also holds for other candidates: e.g., loose couplings of verbal and embodied conduct, locally routinized, or ephemeral gestalts that do not endure beyond the context of their use. My paper contributes to this discussion by proposing a distinction between two kinds of multimodal gestalts: socially sedimented multimodal gestalts (multimodal constructions), and locally assembled, ephemeral multimodal gestalts. To this end, I examine sedimented couplings of demonstratives and embodied practices in instructions, and the change of a locally assembled format over time. The data are in German and come from 12 h of video-recordings of self-defense trainings for young women. In the course of the participants' interactional history, the multimodal format of the participants' actions changes. The changes concern formal and functional aspects of the resources used to accomplish those actions, their multimodal orchestration, and the temporality of their delivery. The paper makes four claims: 1. In their primordial use in co-present interaction, demonstratives are coupled with embodied practices and request addressees' attention to the speaker's body, i.e., they are tightly and intercorporeally coupled with the embodied conduct of the participants; 2. gesturally used demonstratives are socially sedimented multimodal gestalts, i.e., multimodal constructions; 3. multimodal gestalts may be subject to transformations in the course of multiple repetitions; 4. in my data, the transformations lead to the emergence of a new, reduced format, which, while being locally routinized, is neither grammatical nor grammaticalized.

**Keywords:** demonstratives, embodied demonstrations, multimodal gestalts, routinization, sedimentation, emergence, grammaticalization

## INTRODUCTION

Spoken language and embodied practices have been studied in Conversation Analysis (Streeck et al., 2011; Stivers and Sidnell, 2012) and Interactional Linguistics (Selting and Couper-Kuhlen, 2001; Couper-Kuhlen and Selting, 2018) for decades. These closely related approaches have furthered our understanding of language as a fundamentally temporal phenomenon that adapts to, incorporates, and structurally reflects the dialogical, dynamic, and flexible nature of social interaction. Empirical studies within those frameworks provide evidence for on-line language production and understanding (Auer, 2009a), and to the incremental nature of grammatical and conversational structures (Deppermann and Günthner, 2015). Research on multimodality (Streeck et al., 2011; Deppermann and Streeck, 2018) has integrated the body in studying the temporality of language-in-interaction; it has also begun to investigate the local emergence of grammar-body-gestalts (Keevallik, 2015, 2018a, 2018b) and the change of embodied practices over time (Streeck, 2021).

Conversation-analytic and interaction-linguistic approaches resonate with Emergent Grammar (Hopper, 1987, 2011), a linguistic paradigm originally developed in the context of grammaticalization (Hopper and Traugott, 2003). Grammaticalization research is interested in the emergence of grammatical structures in diachrony. In contrast to grammaticalization's focus on relatively stable grammatical structures, Emergent Grammar argues that grammar is never fixed or stable but is constantly evolving (Hopper, 2015). In the unfinished process of grammar as *emergent*, grammar is not prior to, but an epiphenomenon of verbal interaction and ongoingly reshaped by it (Hopper, 2011, 2015).

Interaction-linguistic work (Streeck, 1995, 2009; Auer, 2009b; Stukenbrock, 2018a) provides evidence for homologies between grammar and interaction, in particular, between action projection and grammatical projection (Auer, 2005). These homologies are grounded in the temporal, online quality of grammar (Auer, 2009a; Hopper, 2015), suggesting a close relationship between grammar and interaction. Grammar can be seen "as the historical result of sedimentation and (partly normative) regularization of certain interactional projection techniques" (Auer, 2005: 33).

Interactional Linguistics explicitly "recognizes the effects of past linguistic development, with its sedimentations and ritualizations, and of social historical institutionalization" (Couper-Kuhlen and Selting, 2018: 542). An important characteristic in the interaction linguistic conception of grammar is therefore the sedimentation of a structure in time and social space. Recently, the claim has been made that grammar also emerges from embodied conduct (Keevallik, 2018a). This has stimulated a discussion in Conversation Analysis (CA) and Interactional Linguistics (IL) whether the routinization of embodied practices can be described in terms of grammar and grammaticalization, in other words, whether "grammaticalization and bodily action [...] go together" (Couper-Kuhlen, 2018; Streeck, 2018).

The aim of my paper is to contribute to the discussion on grammar and the body by proposing a distinction between grammar-body constructions and ephemeral grammar-body gestalts, i.e., local, *ad hoc* assembled multimodal gestalts. To this end, I first investigate widely used, socially sedimented grammar-body constructions: couplings of demonstratives and embodied practices. I argue that these constitute prime examples of multimodal constructions (Stukenbrock, 2010, 2015; Ningelgen and Auer, 2017) as part of grammar. They are grammaticalized ready-mades that language communities "inherit" from their ancestors. Second, I examine an *ad hoc* assembled multimodal gestalt and show how it changes in the course of multiple repetitions. As a locally routinized multimodal gestalt, it is not sedimented beyond the ephemeral context of its use and is therefore not grammaticalized. The data are in German and come from video-recorded self-defense trainings for young women.

My paper is structured as follows: In the following section (*Grammaticalization and embodied action*), I discuss the central concepts that bear on my endeavor. Next, *Data and Methodology* are presented. In the first part of the analysis (*Sedimented multimodal constructions as resources in social interaction*), I analyze how grammar-body constructions ("so"/"like this" + gaze + embodied practices) are locally mobilized in social interaction: First, I focus on how gaze projects the focal space for an embodied action. Second, I investigate how the focal moment of bodily performance is indexed by "so"/"like this". Third, I show that couplings of demonstratives and embodied practices form sedimented, yet temporally variable and flexible multimodal constructions. In contrast to the first part of the analysis, the second part investigates a locally assembled, ephemeral multimodal gestalt and tracks its formal and functional change through multiple repetitions: I set out with an analysis of the most elaborate format and subsequently show how the first repetition already exhibits reduction. Next, I illustrate that an increase in complexity indexes and reflects additions or changes in the speaker's utterance. Last, I examine how the format changes in the course of multiple repetitions and undergoes significant reductions. These emerge from routinization and promote automatization as discussed in the concluding section.

I put forward the following claims: 1. In their primordial use in co-present interaction, demonstratives are coupled with embodied practices and request addressees' attention to the speaker's body (Stukenbrock, 2018a; 2018b; 2020a), i.e., they are tightly and intercorporeally coupled with the embodied conduct of the participants; 2. gesturally used demonstratives constitute socially sedimented multimodal gestalts, i.e., multimodal constructions; 3. multimodal gestalts (both grammaticalized or locally assembled) may be subject to transformations in the course of multiple repetitions; 4. in my data, these transformations lead to the emergence of a new, reduced format, which, while being locally routinized, is neither grammatical nor grammaticalized (Hopper and Traugott 2003).

## GRAMMATICALIZATION AND EMBODIED ACTION: (WHEN AND HOW) DO THEY GO TOGETHER?

The term *grammaticalization* refers to “the change whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions and, once grammaticalized, continue to develop new grammatical functions” (Hopper and Traugott, 2003: XV). In the process, their meaning becomes more general and abstract; they fit a broader range of contexts and increase in frequency. Generalization, change in distribution and increase in frequency are mutually reinforcing processes, since generalization facilitates use in more and varied contexts, which then also increases the frequency of the structure (Bybee, 2014: 157). Two perspectives are broadly distinguished: The diachronic perspective focuses on the sources and steps that linguistic structures undergo in the process of grammaticalization; in contrast, the synchronic perspective views grammaticalization as “a syntactic, discourse pragmatic phenomenon, to be studied from the point of view of fluid patterns of language use” (Hopper and Traugott, 2003: 2).

Grammaticalization holds that grammar “is not a static, closed, or self-contained system, but [that it] is highly susceptible to change and highly affected by language use” (Bybee, 2014: 145). The theory of *Emergent Grammar*, which was originally developed within the grammaticalization framework, goes much further and deconstructs the concept of grammar as a system altogether. This is expressed in the term *emergent*. It refers “to the fact that a grammatical structure is always temporary and ephemeral” (Hopper, 2011: 26), and that grammatical forms never become fixed or stable. In contrast, the term *emerging* refers to the traditional view of grammar as “a stable system of rules and structures, which may ‘emerge’ (i.e., come into existence) out of a less uniform mix” (Hopper, 2011: 28).

Endeavors to adapt (Auer and Pfänder, 2011a; Pekarek Doehler, 2021; Pekarek Doehler and Balamán, 2021) and extend (Ford and Fox, 2015) Emergent Grammar (Hopper, 1987, 2011) to examine both grammar-in-interaction as well as gestures (Streck, 2021) document the fruitful synergies between Emergent Grammar and CA/IL. All three share the premise that the linear progression along the timeline (Hopper 2015: 252) is fundamental for our understanding of language and grammar. In a recent study on the local emergence of an ephemeral grammatical practice through reuse, Ford and Fox suggest “a cline between ephemerality and sedimentation” (Ford and Fox, 2015: 96). Although the practice does not “survive” the situation of its creation, and therefore does not move further towards sedimentation or grammaticalization, it is “an ephemeral, temporally specific, manifestation of emergence in grammar” that represents “diachrony at its micro-level” (Ford and Fox, 2015: 115). The authors propose a continuum in Emergent Grammar with a radically ephemeral pole and a sedimented pole at each end. Phenomena of Ephemeral Grammar are located at the far evanescent end of the continuum (Ford and Fox, 2015: 97). If we assume that

phenomena of ephemeral grammar exhibit micro-level diachrony and routinization, how do we conceptualize phenomena on historical time scales, i.e., linguistic structures that emerge from routinization over decades and centuries, acquire high frequency and vast, context-independent distribution?

Key terms such as *habituation*, *routinization*, *automatization*, and *sedimentation* are used both in grammaticalization and in CA/IL. Grammaticalization researchers agree that grammaticalization is a form of *ritualization* (Haiman, 1994) or *routinization* (Hopper and Traugott, 2003). Bybee (2014: 153) defines it as a “process of automatization of frequently occurring sequences of linguistic elements” (cf. also Haiman, 1994). Automatization leads to repackaging of formerly separate units, which lose their identity, undergo formal reduction and semantic bleaching (Hopper and Traugott, 2003; Bybee, 2014). Bleaching, or generalization, is *habituation* to repeated items (Bybee, 2014: 157; Haiman, 1994). Habituation arises from “a decline in the tendency to respond to stimuli that have become familiar” (Haiman, 1994: 7); it is an effect of repetition. In short, grammatical items or constructions “are automated, conventionalized units” (Bybee, 2014: 157). Parallels have been drawn with non-linguistic habituation, ritualization, and automatization (Haiman, 1994; Bybee, 2014). These may be viewed as analogous (Haiman, 1994) or parallel (Streck, 2021) rather than similar processes.

In this paper, I use the terms as follows. *Routinization* occurs through repetition; it is accomplished by the individual through reiterated actions and practices. *Sedimentation* is the social and socially shared outcome of jointly or collectively repeating and routinizing verbal and embodied practices. I distinguish between *joint routinization* and *collective routinization*. *Joint routinization* concerns participants engaged in a shared participation framework; they are mutually aware of one another and repeat certain practices and actions. An example would be dance classes (Keevallik, 2015). The encounters may take place face to face (Deppermann, 2018a, c; Deppermann and Schmidt, 2021) as well as in technically mediated or virtual environments (Pekarek Doehler and Balamán, 2021). Joint routinization may lead to local sedimentation within single encounters (Stukenbrock, 2020b) and across participants’ interactional histories (Deppermann, 2018a; Deppermann and Schmidt, 2021; Pekarek Doehler and Balamán, 2021). In contrast, *collective routinization* emerges across time and space among social groups whose members are not mutually aware of one another. An example would be generic uses of personal pronouns among groups of speakers who converge on this use without knowing that they do (Laberge and Sankoff, 1979; Auer and Stukenbrock, 2018). This may in the long run promote grammaticalization. I propose the term *collective routinization* as a heuristic to bridge the gap between micro-diachrony (Ford and Fox, 2015) and *longue durée*, or macro-diachronic, phenomena classically studied in grammaticalization (Hopper and Traugott, 2003). As long as a format or structure remains a local phenomenon, it is not grammaticalized. For a format to be grammaticalized, it has to spread beyond the initial context of its use, expand and generalize across types of contexts (Hopper and

Traugott, 2003; Bybee, 2014) until it becomes widely used in the language community. This is the case with demonstratives. In the course of *longue durée* processes, they emerged as language universals (Diessel, 1999, 2006; Diessel and Coventry, 2020) and were intricately connected to concurrent uses of embodied attention directing devices such as gestures (Bühler, 1990[1934]). Gestures are an integral component of demonstratives in their primordial, exophoric use in face-to-face interaction. They are part and parcel of the grammaticalized format of demonstratives. Couplings of demonstratives and gestures are grammaticalized ready-mades that members of language communities ‘inherit’ from their ancestors. This contrasts with the reduction and routinization of an *ad hoc* assembled multimodal gestalt. As the analysis will show, its transformation in the course of multiple repetitions indexically reflects and actively promotes routinization of the practices involved: routinization (and even automatization) of motor skills through repetition of self-defense practices; second, routinization of communicative practices through repetition of instructions.

## DATA AND METHODOLOGY

The paper proposes a distinction between two kinds of multimodal gestalts: grammar-body constructions and ephemeral grammar-body assemblages. To contrast usages of a grammaticalized multimodal construction (*so* “like this” + embodied practices) with the emergence of an ephemeral multimodal assemblage, I track the occurrence of their uses in a series of embodied instructions delivered in self-defense trainings.

Instructions have been investigated in a range of settings such as driving (De Stefani and Gazin, 2014; Deppermann, 2018a, b, c; Rauniomaa et al., 2018), air traffic control training (Arminen et al., 2014), cooking (Mondada, 2014a), medical interaction (Svensson et al., 2009; Mondada, 2014b), class room interaction (Lerner, 1995; Lindwall et al., 2015), teaching and learning of bodily skills (Lindwall and Ekström, 2012; Stukenbrock, 2014; Keevallik, 2015; Evans and Lindwall, 2020). The focus has been on how embodied actions figure in the sequential and temporal organization of first and second action (Lindwall and Ekström, 2012; Stukenbrock, 2014; Keevallik, 2015), on multimodal practices of turn construction (Keevallik, 2015), and on changes of turn design over interactional histories (Deppermann, 2018a). Most relevant for my own interest in routinization and reduction are Deppermann’s findings: Within the framework of interactional histories between driving instructor and student, instructions become increasingly shorter, syntactically less complex, and sequentially more condensed. A similar development will be observable in my data.

My study is based on 12 h of video material of self-defense trainings for young women. The participants followed the training voluntarily in their free time. Ethical review and approval were not required for this study. Informed consent was obtained from all participants. The data were recorded with a single, high-resolution video camera and imported into ELAN for verbal transcription and multimodal annotation. All data, including images of the participants, were anonymized. The

images were transformed into drawings with the help of the program *Tayasui Sketches* (<https://tayasui.com/sketches/>).

The data were recorded in different gyms with a focus on the trainer. Around 25 students participated in the classes. They had no previous experience with self-defense trainings. Apart from the trainer and the trainees, one or two student assistants regularly participated to help the trainer arrange materials such as gymnastic mats. In later sessions, they were recruited by the trainer as a partner to enact movement combinations in simulated encounters between victim and aggressor.

For this paper, only the recordings of the initial lessons were taken into consideration. The trainer introduced basic self-defense techniques that were first practiced on their own and then combined to form an embodied whole in the course of the first lesson. A longitudinal perspective across sessions is reserved for a follow-up study on how elements that are already part of the common ground are taken up in subsequent training sessions.

The following analysis is concerned with instructions that refer to self-defense techniques in shared training phases. Instructions that deal with organizational issues were not taken into account. Only cases were investigated in which instructing actions were 1) directed at the whole group and 2) designed to be followed by a performance of the instructed action.

## PART I: SEDIMENTED MULTIMODAL CONSTRUCTIONS AS RESOURCES IN SOCIAL INTERACTION

The focus of the analysis in part I is on the grammar-body construction grounded on the demonstrative “so”/“like this”. It will be shown how embodied demonstrations of the trainer are indexed by the demonstrative “so” and locally designed to fit the addressees’ activities. Progressively assembling a set of resources to mark, co-index and thus emphasize significant moments of embodied actions creates multimodal densifications (“multimodale Verdichtung”, Stukenbrock, 2008, 2015). Multimodal densifications arise from micro-projections at the beginning of an open gestalt and the fulfillment of those micro-projections within that gestalt. The term *gestalt* has been used in multimodal CA for more than 20 years, most prominently in the works of Goodwin (2003, 2007), Heath (1986), Streeck (1988) and others (Streeck et al., 2011; De Stefani, 2014; Deppermann, 2015; Mondada, 2015, 2016; Deppermann and Streeck, 2018). It has been deployed alongside other expressions such as *multimodal packages* or *action packages* (Heath, 1986; Goodwin, 2003, 2007; Streeck, 1995, 2009). Multimodal gestalts are considered to be evanescent phenomena (Mondada, 2015). As such, they resemble phenomena of Ephemeral Grammar (Ford and Fox, 2015). However, couplings of demonstratives and embodied practices are not at the ephemeral end of the “Emergent Grammar-continuum” (Hopper, 2011; Ford and Fox, 2015). Rather, they are prime candidates to argue for multimodal constructions not as locally routinized phenomena, but as sedimented multimodal constructions. They have

grammaticalized the context-bound conditions of their use—this includes, first and foremost, embodied practices (Bühler, 1990[1934]; Stukenbrock, 2015) to establish joint attention (Diessel, 1999, 2006).

The analysis in the first part aims to show how a multimodal construction is deployed in social interaction. The analysis attests to stability as well as to the context-sensitive, temporal flexibility of the construction. It focuses on two components: 1. gaze as a resource to project the focal space for embodied demonstrations, 2. demonstratives as a resource to index the focal moment of an embodied demonstration and, therefore, as a request for gaze.

The couplings investigated in part I are evanescent in real time in situated social interaction. Nonetheless, they are robustly anchored in the language community’s linguistic knowledge via the demonstrative. Demonstratives have grammaticalized our bodily experience with, and joint attention to phenomena in shared space (Diessel and Coventry, 2020; Stukenbrock, 2015, 2020a).

### Projecting the Focal Space for Embodied Action by Gaze

The first extract<sup>1</sup> (“short like this”) shows the beginning of the first self-defense training. The trainer has announced that the students will learn how to mobilize their voice and bodies to protect the territory of the self (Goffman, 1971) against potential aggressors. She decomposes the task into smaller sub-units that are later integrated. We join the group in the course of the first instruction. It is about learning how to make a step forward. The starting point is to stand firmly on the ground. The instruction is addressed at the whole group. In order to be visible to all of them, the trainer has moved to the middle of the gym. The students are arranged around her in full-circle.

The instructional sequence consists of the trainer’s instructing action (l. 1–4) as first pair part (FPP), followed by the instructed action (l. 5) as embodied second pair part (SPP). It is brought to a close by the trainer’s ratification (l. 6) in third position. The trainer’s instruction is delivered as a multi-unit turn. Syntactically, it is built as a conditional construction: The protasis (l. 1–2) formulates and bodily demonstrates the conditions under which the embodied action formulated and performed in the apodosis (l. 4) should be followed. For now, we focus on the multimodal delivery of the first turn constructional unit (TCU), the protasis of the conditional construction. It syntactically projects, first, a subordinate clause that is dependent on the predicate (l. 1: “MERKT”/“realize”, and second, the apodosis.

<sup>1</sup>For reasons of space, only one example is shown in this section. Examples of the grammar-body construction with *so* can be found in the literature (Ningelgen and Auer, 2017; Stukenbrock, 2010, 2014, 2015). Current research on demonstratives provides further evidence for embodiment as part of grammaticalization (Diessel and Coventry, 2020).

```

1 T  also wenn ihr jetzt *MERKT,+
      so when you now realize
      t  -gz to S-----*to floor----->
      s                                     +head down->
2 T  =ihr steht so_KURZ da:,
      you are standing there short like this
3    *(0.7)
      t  *-gz to S-->*
4 T  *dann macht* Einfach n_bIsschen den SCHRITT länger?
      then just make the step a bit longer
      t  *-gz to floor*-to S----->
5 S  (0.2) ((reposition their feet))
6 T  geNAU,
      right
    
```

**EXTRACT 1 |** “Short Like This”

Our analysis focuses on the successive mobilization of linguistic and embodied resources that the trainer uses to project and highlight focal elements of her instruction. The first important moment occurs at the end of the first intonation phrase when the trainer projects a change in the attentional focus by shifting her gaze from the addressees (Figure 1A) to her feet (Figure 1B). Her gaze points to a new space, invites attention-sharing and projects an embodied activity within that focal space.

Extract 1 is a prime example of how embodied demonstrations are integrated into an unfolding verbal instruction. It demonstrates a key function of gaze in conjunction with modal demonstratives (*so* / “like this”) and embodied demonstrations. It projects a new space for embodied demonstrations indexed by *so*. Note that in the extract, the gaze shift precedes the demonstrative, which only comes at l. 2 (see transcript above). As a visible display of human vision, eye-gaze shifts publicly document changes in the attentional focus. In the present case, the gaze shift (l. 1) points to and projects the relevant space for the upcoming demonstration. Before the trainer delivers the demonstrative (l. 2), she thus invites her addressees to follow her line of regard (Stukenbrock, 2020a) and

```

A
B
1 T  also wenn ihr jetzt *MERKT,
      so when you now realize
      t  -gz to S-----*to floor----->
    
```

**FIGURE 1 |** Speaker gaze shift from addressees to focal space.

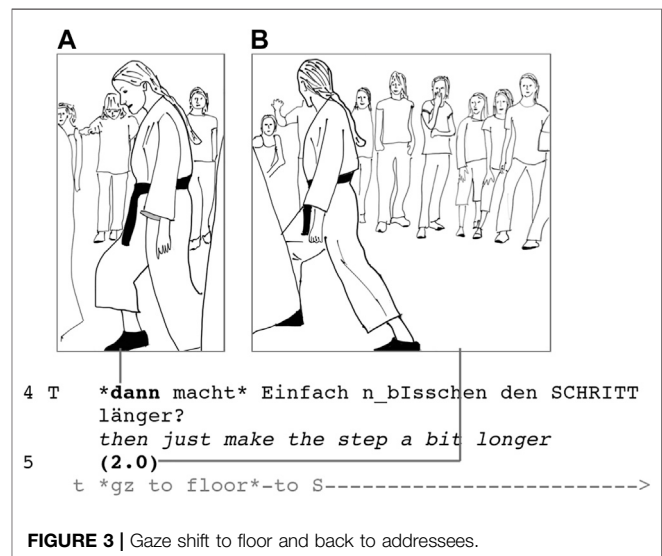
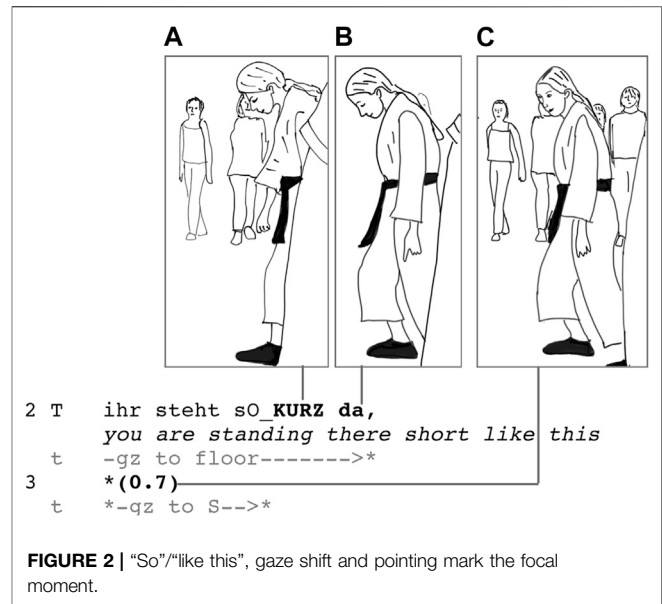
to orient to where the action is going to be.<sup>2</sup> In sum, gaze orientation prepares the **focal space** for an embodied demonstration. As we will see in the next section, the trainer also **temporally** marks the **focal moment** of the unfolding demonstration.

### Marking the focal moment of the bodily performance with “so”/“like this”.

After the trainer has gaze-projected the focal space for the upcoming demonstration (extract 1), she uses the modal demonstrative “so”/“like this” to index the focal moment and element of her demonstration. The demonstrative is part of the second TCU and precedes an adverbially used adjective (l. 2): “ihr steht so: KURZ da”/“you are standing there short like this”. The demonstrative “so”/“like this” is deployed in different constructions (Stukenbrock, 2010, 2015) to index the manner of an action (*so* + VERB), the quality of an object (*so* + presentative constructions), or the degree to which an attributed quality (*so* + ADJ./ADV.) applies to a phenomenon (Stukenbrock, 2010, 2015). It is also used in type-indicative referential actions in conjunction with a noun phrase and a concurrent pointing gesture (Balantani, 2021). It is to be distinguished from uses as a discourse marker (Barske and Golato, 2010), a quotative (Golato, 2000), and various other functions (cf. Stukenbrock, 2014, for an overview). In our example, the demonstrative *so* informs the addressees that the local meaning of the gradable adjective “KURZ”/“short” is to be gathered from the trainer’s embodied action. In temporal terms, it indexes the moment in which the trainer repositions her foot (Figure 2A) to reduce the space between her feet.

Grammatically, “so”/“like this” marks the informational focus of utterance and embodied demonstration; it thus “incorporate[s] the work of the [feet] into the grammatical structure of the talk” (Streeck, 2002: 582). A moment later, the trainer also mobilizes a gesture to point to the space between her feet (Figure 2B). Gaze, demonstrative, body movement, and pointing gesture all work together to highlight (Goodwin, 1994: 606) the crucial moment of her demonstration. Before the trainer continues the syntactic construction (i.e., the projected apodosis of the conditional construction), a pause ensues (l. 3). With frozen body posture, the trainer shifts gaze to the students to monitor their attention (Figure 2C).

At the beginning of the next TCU (the apodosis, l. 4), the trainer shifts gaze once more to her feet (Figure 3A) thus projecting another embodied action to come. The students engage in self-monitoring by looking down at their feet to assess their own spatial position. While describing the corrective body movement that deals with the problematic position demonstrated before, the trainer makes a step forward, and then reorients her gaze to monitor her students (Figure 3B).



By following the trainer’s example and correcting their position (l. 5), the students deliver an embodied display of understanding, which is ratified by the trainer (l. 6: “geNAU”/“right”).

### Sedimented Multimodal Constructions and Temporal Flexibility

In the data, we find temporally variable orders in which demonstratives, gaze shift, and embodied demonstration are mobilized in the local context. Temporal flexibility is not counter-evidence against the claim that couplings of demonstratives and embodied practices are contextually independent, multimodal constructions. On the contrary, flexibility has been from the outset an interactional prerequisite without which the core function of demonstratives would not have emerged: to establish joint attention on phenomena in the shared surroundings of copresent participants. The cross-context

<sup>2</sup>Although the video data do not allow precise observations of the students’ gaze directions, those who are visible at that moment can be seen to slightly accommodate their head orientation downwards.

distribution (Ningelgen and Auer, 2017; Stukenbrock, 2010, 2014, 2015) of these temporally flexible, yet firmly established multimodal constructions has emerged from, and fueled the process of grammaticalization out of which demonstratives emerged as a unique class in linguistic history (Diessel, 2006; 2009; Diessel and Coventry, 2020).

The following extract exemplifies how temporal flexibility allows for variations within the multimodal construction. It documents a local, recipient-designed temporal ordering of gaze, modal demonstrative, and bodily action. It is delivered with respect to the participants' attention and activities. As in extract 1, "SO"/"like this" is coupled with embodied demonstrations and speaker gaze shift from the addressees to the floor. The gaze shift indexes a new focal space to attend to. However, unlike in extract 1, gaze, demonstrative, and bodily demonstration are mobilized in a different temporal order. The trainer shifts her gaze only after the first delivery of the demonstrative, and concurrent with its repetition (l. 3). The trainer's body posture is already in place before the extract starts. She has remained in the stepping position that she assumed before and upholds it throughout the instruction.

```

1 T  ihr sollt jetzt UNgefähr-
    you must now about
    t  -stepping position, gz to S---->
    (0.6)
2
3 T  SO WEIT die füße *SO aus*einander sein,
    wide like this the feet to be apart like this
    t  -gz to S-----*gz to floor*-gz to S-->
3 T  wie eure SCHULtern? (-)
    like your shoulders
4 T  oder eure HÜFten breit sind.
    or as wide as your hips

```

**EXTRACT 2** | "the feet apart like this"

The trainer starts a new instruction with a modal deontic (l. 1: "ihr sollt"/"you must"), moves her arms back and forth along her body, but then breaks off and pauses (l. 2) as some students are still involved in the previous exercise. She restarts with the modal demonstrative "SO"/"like this", which is followed by a gradable adjective ("WEIT"/"wide", l. 3). Instead of projecting a new space of attention by visibly reorienting her gaze to it, the trainer continues to monitor her addressees (Figure 4). Since some of the students are not looking at her, the gaze shift would not be seen and hence interactionally useless.

Up to this point, the demonstrative, instead of being preceded by a gaze shift, precedes the gaze shift. By this temporal ordering, the (first use of the) demonstrative serves as an audible request for addressee gaze (Stukenbrock, 2018b) at a moment when focused interaction and visual coorientation need to be re-established. The demonstrative hearably indexes that visible information is to be gathered from the trainer's embodied action. In order to understand the local meaning of "SO" with respect to the gradable adjective "WEIT"/"wide like this", the addressees will have to look at the trainer.

After the first, multimodally "lean" occurrence of the demonstrative, the trainer shifts gaze from the students to the floor and performs two gestures to delineate the space projected by her body (Figure 5A). Concurrent with her embodied actions, she repeats the modal demonstrative "SO"/"like this" (l. 2),



3 T SO WEIT die füße  
wide like this the feet

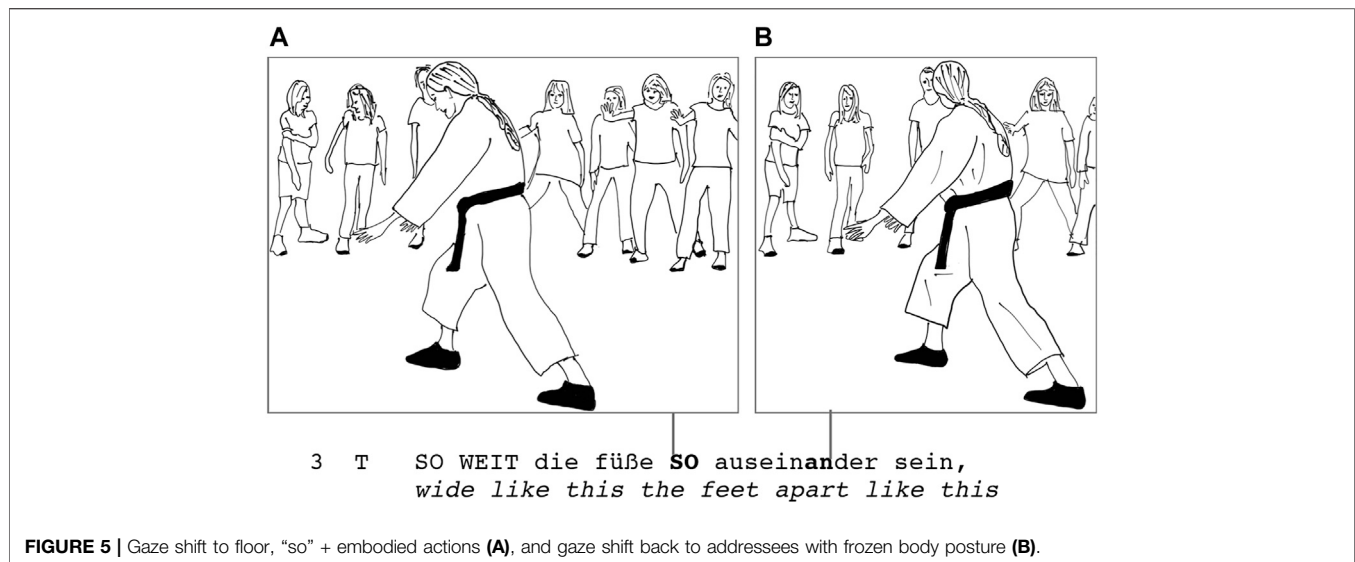
**FIGURE 4** | Continuous gaze at addressees.

freezes her body posture, and shifts gaze back to the students to monitor their attention (Figure 5B).

In contrast to the first extract, where the trainer's gaze shift to a new domain preceded demonstrative and embodied action, it is now the demonstrative (its first delivery) that precedes the gaze shift to the new domain: It implements a summons for addressee gaze (Stukenbrock, 2018b). This use is made contingent on the trainer's perception that some students are still engaged in finishing the previous exercise and not yet ready to look at her.

The extract documents that the resources are recipient-designed to fit the addressees' situated activities. Thus, while the resources (first and second use of modal demonstrative, embodied demonstration, gaze shift) are temporally calibrated to the addressees' diverging foci of attention, they are still converging to "embody" the same kind of multimodal construction. The first, "lean" delivery of the format, which requested visual attention from unattending participants, is followed by a full multimodal delivery of the grammar-body construction in the course of the trainer's self-repair.

To sum up, the analysis in part I has shown that modal demonstratives ("so"/"like this") are closely coupled with embodied actions. These constitute indispensable components without which the demonstrative would not be understood. The speaker's embodied actions have to be seen by the addressees in order for them to understand the local, indexical meaning of the demonstrative. Participants orient towards this need as a joint endeavor: The trainer designs and times her actions with respect to the addressees' attention and availability. Evidence for this was given in extract 2, where the trainer deployed a modal demonstrative to summon the visual attention of non-attending addressees before she recycled the demonstrative as part of a full-fledged multimodal construction. Conversely, addressees



consistently orient to exophorically used demonstratives as requests for visual attention by allocating their gaze to the speaker and attending to her embodied actions.

By default, requests for gaze are formulated by perceptual imperatives. However, they are also delivered by less specialized means, such as restarts and pauses (Goodwin, 1980), prospective indexicals (Goodwin, 1996), response cries (Goffman, 1981), noticings (Keisanen, 2012; Stukenbrock and Dao, 2019), and by combinations of those means (Goodwin and Goodwin, 2012). As we have seen, summons for gaze are also implemented by demonstratives. What is more, this is constitutive for the primordial function of demonstratives in phylo- and ontogenesis. The gaze-summoning property of demonstratives is inherently linked to speakers’ embodied actions and to the need of addressees to perceive those actions. Demonstratives are therefore “by nature” embodied—i.e., multimodal constructions (Ningelgen and Auer, 2017; Stukenbrock, 2010, 2017, 2018a, 2020a).

## PART II: LOCALLY ASSEMBLED MULTIMODAL GESTALTS

I have argued that the multimodal couplings examined in part I are systematic and acquired as part of grammatical knowledge; they underwent grammaticalization long ago and constitute multimodal constructions. In part II, I will investigate multiple repetitions of a multimodal format in the course of the participants’ interactional history. Repetitions are crucial for the emergence of grammar: “Grammar is nothing other (and nothing “deeper”) than repeated and automated motor action, and the best moment to study its emergence, as it were, is the first repetition” (Streeck, 2018: 31). However, there are important differences between the local routinization of ephemeral phenomena and grammaticalization as a *long durée*-process (Streeck, 2018, 2021); the latter transcends particular participation frameworks, local communities of practice, generations, and even centuries. The grammar-body-gestalts investigated in this section are locally routinized. *Via* repetition,

they are sedimented within and for that group. Concurrently, the format becomes increasingly reduced.

### The Elaborate Format

We begin with the most elaborate format and subsequently examine how the format is becoming leaner over time as components are gradually being abandoned. It consists of a request ‘to X something “like this” + gaze to focal space + embodied demonstration’. Extract 3 shows the full format. The trainer requests the students to place their hands on their hips in a particular way. The instructional action (l. 1–2) is followed by an instructed action (l. 3) delivered by students. The sequence is closed as the trainer comments on the practice in third position (l. 4).

```
1 T un_+dEs*wegen nehmt ihr jetzt +ERST+mal,  
and that's why you now for a start put  
gz to S*-gz to hands----->  
t +...lifts hds to hips.....+stroke+hds at hips->  
2 T die (.) hän+de *SO* in [die Hüfte?  
the hands like this on the hips  
t -gz to hands--*...*-gz to S----->  
t -hds at hips+beat +-hds at hips----->  
3 S [((put hands on hips))  
4 (1.0)  
5 T das machen wir später NICHT mehr;  
we won't do that any more later on
```

**EXTRACT 3 |** “hands like this on the hips” (MM\_B1\_00:15:22).

The instructional action (l. 1–2) is delivered multimodally. At turn-beginning, the trainer is looking at her addressees (Figure 6A). She lifts her hands, bends her head, and visibly shifts gaze to her hands (Figure 6B), thus gaze-flagging (Streeck, 2002) her embodied demonstration as it emerges. She continues to gaze down as she moves her hands to her hips in a palm-away position (Figure 6C).

In the course of the second intonation phrase, which contains the demonstrative “SO”/“like this” (l. 2), the trainer produces a gestural stroke by quickly moving her hands sideways and hitting her hips (l. 1), palms away (Figure 7A). The demonstrative is prosodically



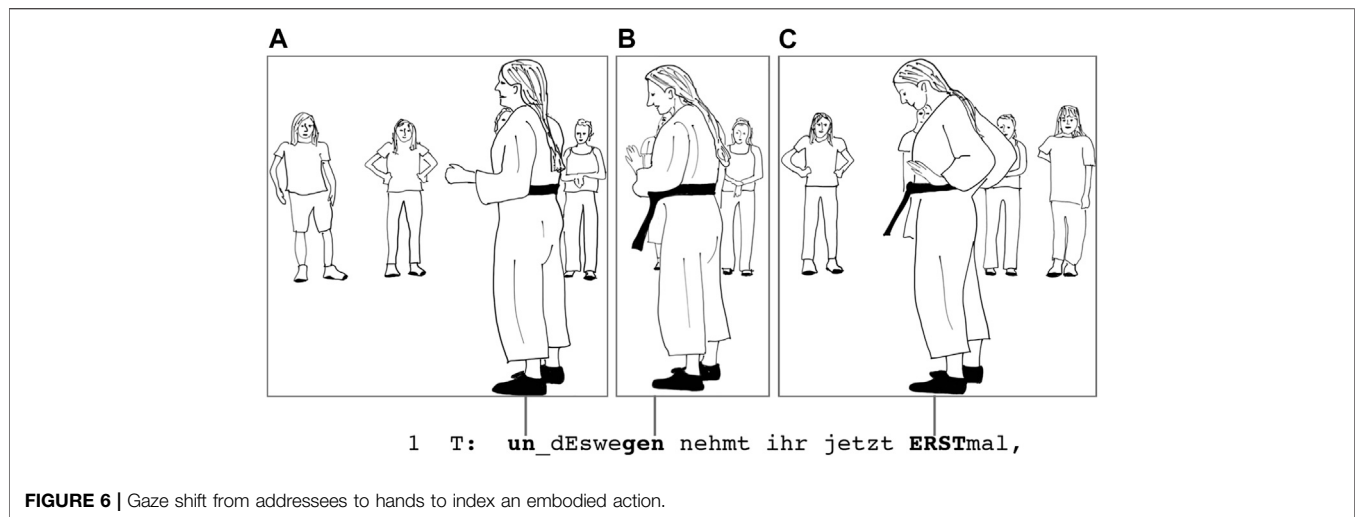


FIGURE 6 | Gaze shift from addressees to hands to index an embodied action.

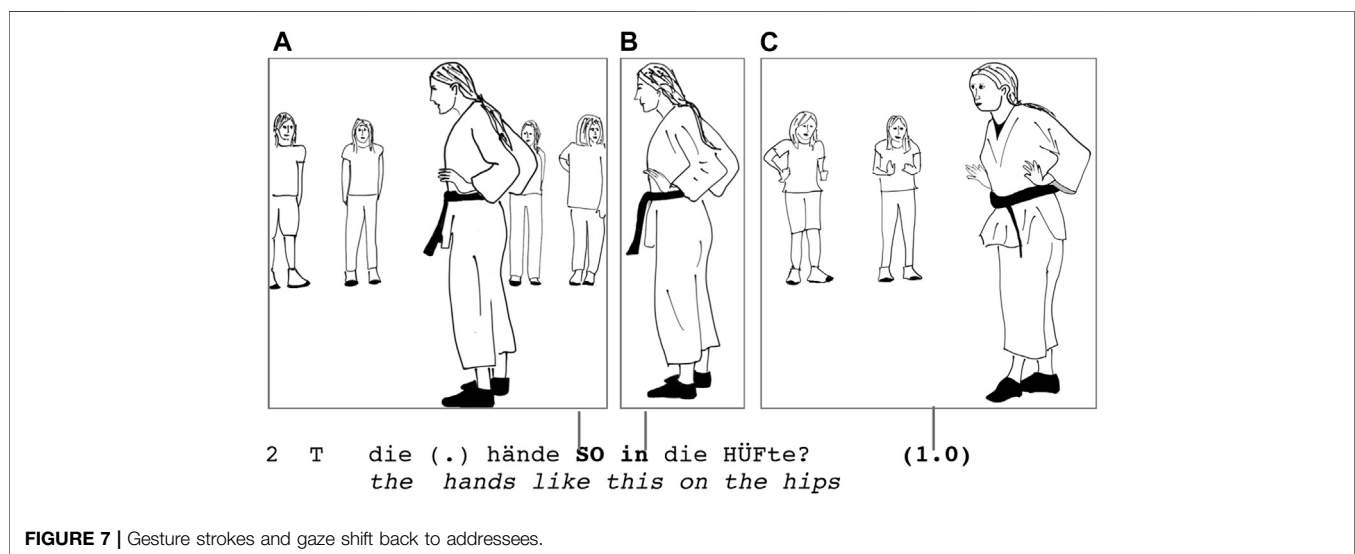


FIGURE 7 | Gesture strokes and gaze shift back to addressees.

marked by a focal accent, and concurrently, the position of the hands is emphasized by a gestural beat, or baton (Kendon, 2004). A second, laterally performed baton occurs concurrently with the delivery of “SO” (l. 2).

After the multimodal gestalt is fulfilled and the turn completed, the trainer shifts gaze to the students (Figure 7B). In conjunction with the high-rising intonation at the end (l. 2), the gaze shift mobilizes an embodied response (Stivers and Rossano, 2010). With a scrutinizing look (Figure 7C), the trainer turns in a semi-circle to check how the students perform the instructed action.

In line with our previous analysis, we can observe a temporally fine-tuned mobilization of resources: While gaze projects the focal space for the embodied performance (cf. *Projecting the focal space for embodied action by gaze*), the demonstrative marks the focal moment of the performance (cf. *Marking the focal moment of the bodily performance with “so”/“like this”*). Gaze, demonstrative, and gestural baton are assembled to co-index, by multimodal densification, the key moment of the trainer’s instruction.

## First Repetition and Reduction

Extract 4 documents the first repetition after the initial instruction in extract 3. Its turn-design differs from that in extract 3, and its multimodal delivery is significantly reduced. First, the trainer has to reorganize the students’ positions and manage the transition to the next round. While the discourse marker *okay* at turn-beginning (l. 1) marks the transition, the organizational instruction “nochmal zuRÜCK”/“back again” realigns the students in interactional space and brings them back to the by now familiar starting position. This is indicated by the temporal adverb “nochmal”/“again” (l. 1). It contrasts with the temporal marker “erstmal”/“for a start” in extract 3, and projects a second go. It is repeated with focal accent as part of the instruction proper (l. 3) and indicates familiarity to the students. The verbal instruction (l. 4) is accompanied by a hands-to-hips-movement and followed by the students’ performance of the instructed action (l. 4).

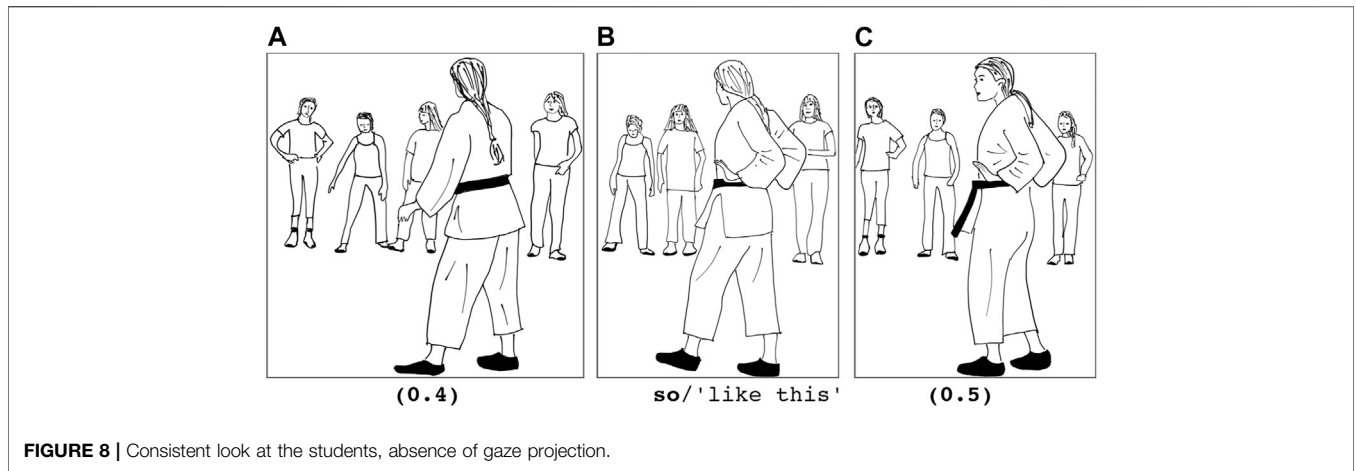


FIGURE 8 | Consistent look at the students, absence of gaze projection.

```

1 T <<acc>okay nochmal zuRÜCK,>
    okay back again
t  -gz to S----->
2 [(0.4)
S [((step back))
3 T <<acc>+noch+MAL die hände so in die hüfte?>
    again the hands like this on the hips
t  +lifts hds+hds at hips palm away---->
4 (0.5)

```

**EXTRACT 4** | “again hands like this on the hips”.

TABLE 1 | Summary of reductions.

First delivery of instructing action (extract 3)	First repetition of instructing action (extract 4)
speaker gaze flag: publicly visible projection of focal space	no speaker gaze flag: interactionally known focal space
demonstrative with focal accent: audible temporal projection of focal moment	no focal accent: interactionally known focal moment
deictic address term and descriptive verb phrase	reduction to truncated deontic infinitive
multi-unit turn, two intonation phrases	single intonation phrase

Before the trainer delivers the instruction, she publicly displays that she is monitoring the students’ activities (l. 2, **Figure 8A**). In contrast to extract 3, where she projected the focal space of the instruction by gaze, she now consistently looks at the students (**Figure 8A,B,C**). By turning her head and visibly letting her gaze wander across the group (**Figure 8C**), she documents that she is closely monitoring the students’ embodied response.

Further reductions are observable: In extract 3, the instructing action was delivered in two intonation phrases (l. 1–2). In contrast, it is compressed into a single one in extract 4 (l. 3). Whereas the trainer used a proposition with a deictic address term (“ihr”/“you”) and an inflected verb phrase (“nehmt”/“take”) in extract 3, she now uses a truncated deontic infinitive instead (on deontic infinitives cf. Deppermann, 2006). Moreover, she omits the gaze shift to the focal space (spatial projection), and downgrades the prosodic design of the demonstrative<sup>3</sup> by shifting the focal accent to the adverb (l. 3: “NOCHmal”/“again”). By repeatedly indexing that the instruction is already part of the common ground, the trainer accounts for a scaled-down version of the instruction: Visibly projecting the focal space by gaze and audibly emphasizing the crucial moment by a prosodically marked demonstrative is less important when these are already known to the participants. The reduction is summarized in **Table 1**.

The short excursus in the next section contrasts our analysis of repetition, routinization, and reduction with the opposite case. When

the trainer introduces new elements, the instruction becomes more complex again. Against this background, the eroding effect of multiple repetitions (cf. sub-section *Local routinization and sedimentation through repetition and reduction*) will become even more apparent. Furthermore, we can also see from the contrasting example how incipient routinization can be stopped or blocked.

### Excursus: Meta-instructions to mark an addition, a change, or a new instruction

In this short excursus, it is argued that while multiple repetitions lead to routinization, simplification, and reduction, the opposite—introducing new elements—motivates the use of extended, more complex formats. The choice and design of the format thus reflexively indexes familiarity and routinization or lack thereof.

The extract occurs after repetitions have already yielded initial reductions. However, it does not exhibit those reductions. On the contrary, it is more complex than the previous extract. The reason for this is that the trainer introduces a new element. She delivers a meta-instruction to announce that element. The meta-instruction establishes a hand clap as a timing signal for choric practicing.

Meta-instructions add a layer of reflexivity to the reflexivity and indexicality of situated social interaction by explicitly formulating an instruction about instructions. They establish local practices of co-orientation and co-ordination, and request attention to and alignment with those practices of practicing. They formulate practices for the local organization of instructions-in-interaction. Relevant for my argument is that meta-instructions, and more

<sup>3</sup>Note that this observation does not question the observation that demonstratives (modal as well as spatial) when used exophorically, and gesturally, in face-to-face interaction, bear the focal accent of the intonation phrase. They do, and only in an uptake or repeated use may the resources, in this case the accent, be reduced.

generally, meta-formulations (re-)increase the complexity of formats that may have begun to undergo reduction.

```

01 T jetzt machen_wa_s SO,
    now we are gonna go like this
02 Ich klatsch in die HÄNde,
    I clap my hands
03 und Ihr geht Einfach diesen schritt nach vOrne? (-)
    and you simply make this step forward
04 ACHtet drauf,
    pay attention
05 dass ihr (--) BREIT genug seid?
    that you are wide enough
06 <<acc>also nIcht wie_n SEILtanz macht,> (-)
    PTCL don't go like tightrope
07 und vOrne im KNIE steht;
    and stand forward in the knee
08 <<acc>und dann GÜCKen wir mal-
    and then we take a look PTCL.
09 was man nOch so alles (.) FALSCH machen kann;
    what else PTCL one can also do wrong
10 UND- ((claps hands))
    and
11 (0.2)[(0.3)
    S [(make a step forward)]
12 T <<p>o*KAY;>
    t *turning and looking around->

```

**EXTRACT 5** | “I clap my hands”

The trainer starts with an announcement (l. 1–3). She uses a pre-construction with “SO”/“like this” (l. 1) (“Vorlaufkonstruktion mit so”, cf. Auer, 2006), which projects a prosodically and syntactically complex turn. The subsequent bipartite turn delivers the meta-instruction (l. 2–3) and fulfills the syntactic projection. It introduces the hand clap as a timing device for choric practicing. While the instructional object (stepping forward) is referred to as already known (l. 3: “diesen schritt”/“this step”), the method of choric practicing according to the hand clap is introduced as something new. It is defined as go-ahead for the students’ performance of the instructed action. In grammatical terms, it functions like a gesturally used temporal demonstrative that points to the moment of its utterance (Fillmore, 1997; Levinson, 2005).

The trainer delays the delivery of the hand clap and thereby holds back the students’ response. She inserts instructional details on how the step forward should (not) be done (l. 4–7), and announces an assessment of trouble sources that the students may be exhibiting in the course of the performance (l. 8–9). The trainer projects and designs an action trajectory that is composed of her hand clap as FPP, the students’ performance as SPP, and subsequently, further assessment and training phases that target the students’ problems as they become visible to the trainer’s professional vision (Goodwin, 1994). By publicly anticipating problems, the trainer prospectively accounts for the need for future correction and repetition.

The trainer performs the hand clap with a large, sweeping movement, which prepares the stage for the audible go-ahead. Additionally, the hand clap is projected by a pre-positioned, prosodically marked verbal item: the conjunction “UND”/“and” (l. 10). The students respond by stepping forward after the hand clap.<sup>4</sup> The trainer acknowledges the performance and prepares the transition to the next round with “oKAY” (l. 12).

<sup>4</sup>Although the clap is projected by other resources, the students never move forward in synchrony with the clap.

The trainer uses the hand clap as a device to structure the instructing action, insert details, anticipate problems, and delay the students’ performance by withholding the clap and making its delivery contingent on the ongoing activities. The sequential structure can be summarized as follows:

- I. position: complex multi-unit turn of the trainer composed of
  - 1) announcement, couched in a pre-construction (“Vorlaufkonstruktion”) with “so”/“like this”
  - 2) meta-instruction to establish trainer’s hand clap as go-ahead for students’ step forward
  - 3) insertion of instructional details
  - 4) preview of further assessment and repetition sequences
  - 5) and-prefaced hand clap as go-ahead
- II. position: students’ embodied response
- III. position: ratification by trainer

The analysis shows that complex, multi-unit turns with pre-positioned announcements and meta-instructions reflexively constitute and index the additional effort to formulate changes in the instructional format. The complex format used to formulate new and unfamiliar elements contrasts with reductions exhibited as the result of repeating the familiar. Multiple repetitions and reductions may ultimately lead to the local emergence of a new format. This is studied in the next sub-section.

## Local Routinization and Sedimentation Through Repetition and Reduction

Previously, we have seen how first repetitions already exhibit reductions. The short excursus on meta-instructions, in contrast, showed how the introduction of new elements leads to increased complexity, which may eventually counteract routinization and reduction. In this sub-section, we study how the complex, multi-unit turn format is once again changed and reduced in the course of multiple repetitions. The analysis focuses on reductions that emerge from progressive routinization of first and second actions, and on the concurrent temporal compression that reflects and constitutes initial automatization.

```

01 T oKAY;
    t -gz at S--->
02 T nochmal zurÜCK?
    back again
03 S ((step back))
04 T und JETZT geht ihr auf_s klatschen mit dem Andern bein vor;
    and now you step forward to the clap with the other leg
05 T UND=((claps hands))
    and
06 S ((step forward))
07 *(0.4)
    t *turns rounds, monitors S-->
08 T oKAY;

```

**EXTRACT 6** | “step forward to the clap”

Extract 6 occurs right after extract 5. It exemplifies how subsequent repetitions allow for further reductions. The reductions concern both the meta-instruction and the instruction proper. The hand clap has already been put to practice as a timing signal and is reused as a go-ahead in the subsequent instruction.

The first reduction concerns the meta-instruction. Whereas in extract 5 it was delivered in a syntactically, prosodically, and pragmatically complete TCU and followed by the instructing action, it is now boiled down to a prepositional phrase (l. 04: “auf\_s klatschen”/“to the clap”) and integrated into the instructing action (l. 04: “und JETZT geht ihr auf\_s klatschen mit dem Andern bein vor”/“and now you step forward to the clap with the other leg”). Although the clap is already established as a go-ahead, the trainer recycles the meta-instruction as part of a modified instruction: She now requests the students to step forward with the other leg (l. 04).

As in extract 5, the trainer projects the hand clap by a prosodically marked *and*-preface (l. 05). In contrast to extract 5, however, she no longer visibly puts the hand clap on stage. Instead, it is latched to the *and*-preface and done very quickly. The students subsequently perform the instructed action, and the sequence is closed when the trainer, after turning around to monitor the students (l. 07), utters a ratification (l. 08: “oKAY;”).

The next extract documents further reductions. Again, the trainer uses a meta-pragmatic announcement, but marks the practice as already familiar by the modal adverb “wieder”/“again” (l. 01). While the practice of clapping and most of the instructed action are treated as known, a new element is introduced: raising the arm when stepping forward (l. 02). In contrast to extract 6 where the announcement of the clap and the instruction were delivered in a single TCU, the trainer now constructs two TCUs and thus foregrounds the arm raise as an instructional novelty.

```
01 T ich KLATSCH wieder,
    I clap again
02 und ihr NEHMT (-) den ARM mit dazu;
    and you use your arm as well
03 T ((claps hands))
04 S ((step forward and raise arms))
05 T oKAY;
06 =nochmal zurÜCK?
    back again
```

**EXTRACT 7** | “Short Like This”

A formal reduction and temporal acceleration occurs in the adjacency pair of trainer’s gestural go-ahead and students’ embodied response (l. 3–4). The trainer now omits the *and*-preface, which formerly projected the hand clap and gave the students time to prepare. Instead, she claps immediately after the delivery of the instruction (l. 3). Subsequently, the students step forward and raise their arms (l. 4). The trainer uses the same item for ratification (l. 5: “oKAY”), but now latches an organizational instruction that projects a next go. A double acceleration is thus accomplished: Omitting the *and*-preface temporally compresses verbal instruction and gestural go-ahead; latching the organizational instruction to the ratification speeds up the succession of training rounds.

The next extract starts with a correction (l. 1) and an organizational instruction (l. 02). It implicates repetition and is marked as part of the interactional history by the temporal adverb “NOCHmal”/“once more” (l. 2).

```
01 T Immer nur EIN ARM;
    always only one arm
02 NOCHmal (.) HINTen in die HÜFte,
    once more back in the hips
03 S ((put arms and hands back near the hips))
04 T U:ND [ZACK;
    and zack
    [((claps hands))
05 S ((step forward and raise arms))
06 T oKAY;
07 wieder zurÜCK?
    back again
```

**EXTRACT 8** | “and + [hand clap + zack]”

For the first time, the trainer now leaves out the meta-pragmatic announcement. Instead, she re-introduces the *and*-preface (l. 04) to project the hand clap, and adds a new element: the vocalization “ZACK” (l. 04: “U:ND ZACK”/“and zack”). The interjection *zack* onomatopoeically indexes a sharp and violent movement (DWDS; GRIMM, Bd. 31, Sp. 10). In the context of self-defense trainings, it not only depicts these movement qualities, but mobilizes the students to perform the instructed action with utmost force and velocity. By synchronizing the delivery of vocalization and hand clap, the trainer performs a very short and sharp go-ahead signal. In contrast to the concise, synchronized delivery of hand clap and vocalization, she lengthens the pre-positioned conjunction “U:ND”/“and” (l. 04). The delay contrasts with and thus highlights the subsequent acceleration of the go-ahead, which invites a fast and forceful response.

This exercise is repeated two more times, with an explanatory sequence in between. The two repetitions are delivered in the reduced format with an *and*-preface to project the clap and a synchronized performance of clap and vocalization as go-ahead for the embodied response.

After the arm raise has been repeated several times, the trainer announces the last element to be integrated. The students will now also have to perform a scream. The scream has been practiced separately before. The trainer delivers the instruction in a more complex format. This choice is in line with our observations in the excursus on the increased complexity and length of instructions that introduce additions or changes.

```
01 T ich KLATSCH wieder?
    I clap again
02 und ihr geht ALLe, (-)
    and you go all of you
03 mit dem !NEIN! SCHREI nach vOrn.
    with the no scream forward
04 U:ND [ZACK;
    and zack
    [((claps hands))
05 S ((step forward, raise arm and scream))
06 T SuPer okAy;
    great okay
07 nochmal zurÜCK?
    back again
```

**EXTRACT 9** | “and + [hand clap + zack]”

First, the trainer returns to the meta-pragmatic announcement (l. 1) that she had left out before; second, she delivers the instruction in two intonation phrases (l. 2–3). These are separated by a small pause.

The syntactic gestalt of the first intonation phrase (l. 2) is incomplete and projects more to come. The pause between first and second intonation phrase becomes hearable as a turn-holding device, which slightly delays the second intonation phrase (l. 3), summons the students' attention, and brings the new instructional component, the "no-scream" (l. 3), into focus. Next, the trainer uses the reduced format of *and*-preface, simultaneous vocalization, and clap (l. 4). After the students have integrated the new element (l. 5), the trainer formulates a positive assessment (l. 6: "SUper"/"great"). Subsequently, she requests the students to step back and projects a repetition (l. 7).

After a brief comment on the scream, the trainer recurs to the lean format. The lean version documents progressive reductions of the instructing FPP and concomitantly, a temporal compression between FPP and embodied SPP.

```
01 T U:ND [ZACK;
      [((claps hands))
02 S ((step forward, raise arm, and scream))
```

**EXTRACT 10** | "Short Like This"

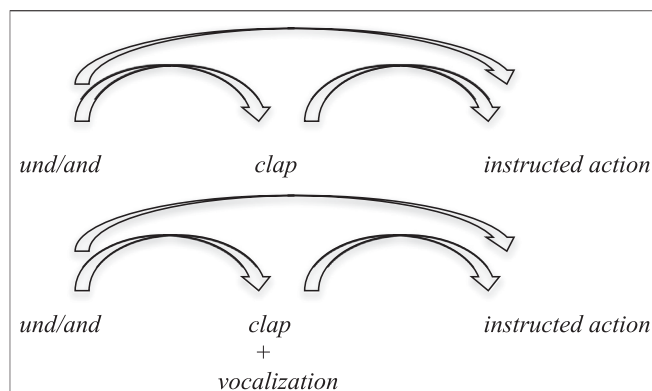
While the *and*-prefaced coupling of vocalization and hand clap projects the temporal slot for the students' embodied response, they have to infer from the interactional history how to design their action. In the present case, they understand the trainer's minimal signal as a go-ahead to repeat the previous action (l. 2). No explicit instruction tells them that they are requested to repeat the integration of the three elements they have practiced separately before.

The next extract attests to the local adaptability and temporal flexibility of the format once it has been established. These variations do not constitute counter-evidence to the observed integrity of the format as an oriented-to, recognizable gestalt. They are recipient-designed temporal calibrations. The local variations reflect and orient to the addressees' attention and participation. The fact that the reduced format can be lengthened without being fragmented is evidence to its beginning sedimentation in the local context. The formal and functional sedimentation of the format is the result of the participants' joint routinization.

```
01 T U:::ND [(0.7) ] [ZACK;
      [((claps hands))
      t --gz at S----->
      S [((laughing))]
02 (0.2)
03 G ((step forward, raise arm and scream))
```

**EXTRACT 11** | "and + [hand clap + zack]"

In extract 10, the trainer's action was designed and understood as a repetition of her action in extract 9. In the same way, her action in extract 11 is delivered and understood as a repetition of her actions in extracts 9 and 10. However, extract 11 exhibits a significant temporal variation: The *and*-preface is extremely lengthened and followed by a long pause before the go-ahead is delivered (l. 01). As some students are laughing among themselves (l. 01), the trainer delays her action by adapting it to the students' activities and attentional focus.



**FIGURE 9** | Projection of timing resources.

Originally, the hand clap was introduced as a go-ahead only, and later coupled with a vocalization. In order to project the occurrence of the go-ahead, the trainer used a prepositioned conjunction, the *and*-preface. The format "*and* + [clap + vocalization]" was used in two sequential contexts: 1) after an instructing action with a new component to initiate and time the students' performance of the new practice, 2) to invite and time repetitions of an established practice. After several alternations between 1) and 2), the format began to index, even after insertions, by inference alone, the most recent practice. In other words, it has progressively assumed the meaning and function of what has been left out, and has finally become a shibboleth for the instructing action.

Although both devices, the *and*-preface and the clap, project and time what comes in the subsequent slot, their function developed along different paths in the course of the participants' interactional history. Whereas the trainer explicitly established the timing function of the clap by a meta-pragmatic announcement, the projecting function of the *and*-preface emerged in practice.<sup>5</sup> It is, moreover, based on the projective properties of syntax in German (Auer, 2015). This can neither be claimed for the clap nor for the vocalization, notwithstanding the fact that they also project what comes next. However, their projective force is grounded in interaction, and not in grammar (Auer, 2005).

The *and*-preface is combined with the clap to form a syntagma of progressively projecting timing resources. The same holds for the vocalization, which was introduced and routinized by practice and which inherited the function of the meta-pragmatically established and synchronously performed clap (Figure 9).

After the format has been repeated several times, the trainer returns to the minimal version even after insertion sequences. She no longer goes back to a more complex format

<sup>5</sup>Questions that emerge from this are, which elements lend themselves to being introduced *en passant*, in and through practice alone, and which elements are, in contrast, metapragmatically established, and why is this so. These problems cannot be discussed here. They are topics for further investigation.

in order to redesign the FPP. This is further evidence to an increased sedimentation of the format “*and* + [clap + vocalization]”.

## When Drill Takes Over to Automatize Motor Actions

The last extract documents that in the course of multiple repetitions, the format undergoes still further reduction. Extreme reduction and acceleration finally transform routinization into automatization. Note that this not only constitutes a qualitative change, but once more raises the question of grammaticalization, if grammaticalization is automatization (Bybee, 2014) and grammar “nothing other than [...] automated motor action” (Streeck, 2018: 31). This question will be discussed in the final section.

Extract 12 shows the maximally reduced format. The trainer now simply claps, and the students subsequently perform the instructed action.

01 T: ((claps hands))  
02 G: ((step forward, raise arm and scream))

EXTRACT 12 | “Clap Only”

This extreme reduction enables an even faster transition between first and second action, between clap and step. At the same time, it significantly accelerates the succession of repetitive goes at the same action. In order to accelerate and automatize students’ motor actions, the trainer progressively shortens her action, accelerates, routinizes, and finally automatizes the temporal succession of FPP and SPP. Training units that are repeated over and over again undergo acceleration, dynamization, and automatization. These features reflect not only a quantitative, but also a qualitative change in the participants’ exercising practice: It is ultimately transformed into drill. Drill as a practice in military and in sports serves routinization and automatization of motor actions performed innumerable times at high velocity.

The phenomena described in this sub-section exhibit striking parallels with processes of grammaticalization (cf. *Grammaticalization and embodied action: (when and how) do they go together?*). On the one hand, the extracts testify to progressive routinization and acceleration of motor actions that the students repeat multiple times **in order to automatize and incarnate them** as part of their repertoire of self-defense techniques. On the other hand, the extracts document a process of routinization, sedimentation, and even automatization that takes place on a different plane: communication. Embodied resources are used and coupled with speech **in order to communicate, to deliver verbal actions**; they are not repeated in order to learn and automatize language—as in old-school language teaching –, but in order to deliver and structure verbal actions, and to project and time addressees’ embodied responses. In the activities under investigation, the latter—reducing and accelerating communicative actions—is in the service of the former—accelerating and routinizing motor actions. These processes are not separate, but intertwined, they reflexively constitute and index co-emerging properties. The

communicative practices used to teach self-defense practices inherit properties of the latter while the latter are shaped by the communicative practices of the former.

## DISCUSSION

The aim of this study was to contribute to the recent grammar-body-debate by proposing a distinction between two kinds of grammar-body-gestalts: 1. socially sedimented, grammaticalized multimodal constructions, and 2. locally routinized ephemeral gestalts. Evidence for the first type was provided in part I of the analysis by an examination of modal demonstratives, embodied practices and concurrent gaze behavior. The focus was on demonstrations indexed by the modal demonstrative *so* / “like this” and “flagged” (Streeck, 2002) by speaker gaze. In line with typological, historical, and interaction linguistic studies on demonstratives, it was argued that the primordial function of demonstratives is to establish joint attention on phenomena in the participants’ surroundings, and that this makes embodied devices indispensable (Bühler, 1990[1934]; Diessel and Coventry, 2020; Stukenbrock, 2020a). Embodied practices are made of participants’ motor actions; these unfold in time, exhibit “inner duration” (“innere Dauer”, Streeck, 2007: 158), and are interpersonally coordinated. Temporal flexibility is therefore an interactional prerequisite without which demonstratives as multimodal constructions could not have emerged. In short, temporal flexibility is the sedimented historic result of concrete, situated, temporally fine-tuned uses of those grammar-body-gestalts in language history.

In social interaction, the use of these constructions is made contingent on the local context, the resources are mobilized, recipient-designed, and temporally calibrated to fit participants’ ongoing activities. In other words, while these constructions are made of emerging (historically sedimented, grammaticalized) constructions; they are delivered in context-sensitive ways as emergent constructions. Here, variation and innovation take place, and new ephemeral multimodal gestalts emerge. When these are reiterated, routinized, and distributed across contexts, they may eventually become grammaticalized.

In part II, I investigated the emergence of such an ephemeral multimodal assemblage and its micro-diachronic changes. It was shown that in the course of multiple repetitions, the multimodal gestalt underwent formal reduction and functional change. Although the observed processes and changes are similar to those described in grammaticalization, radically different temporal scales and social-distributional dimensions are involved. As long as a format or structure remains a local phenomenon, it is not grammaticalized. It has to spread beyond the initial context of its use, expand, and generalize across types of contexts (Hopper and Traugott, 2003; Bybee, 2014) until it becomes widely used in the language community and part of its shared linguistic repertoire or knowledge as an effect of “social historical institutionalization” (Couper-Kuhlen and Selting, 2018: 542).

In sum, multimodal gestalts with different histories are evoked in social interaction. Ephemeral multimodal gestalts are not

grammaticalized and have no place in grammar. I do not claim that locally occurring, ephemeral gestalts cannot be grammaticalized. Rather, my proposition is to distinguish between micro-diachronic and historical processes, and to consider joint routinization and collective routinization as subsequent stages along a path towards grammaticalization. The cradle for such a development may be the movement of a practice from the ephemeral pole to the sedimented pole of the Emergent Grammar-continuum (Ford, and Fox, 2015). But it has to move on beyond the sedimented pole of Emergent Grammar and along the grammaticalization path (Hopper and Traugott, 2003) that leads to social sedimentation and institutionalization across contexts. This view approaches (multimodal) constructions both as emerging and emergent (Auer and Pfänder, 2011b). It emphasizes that “[t]here is no need to exclude routines from an emergentist approach” (Auer and Pfänder, 2011a: 18), and, in turn, that emergent constructions are the stuff that emerging constructions are made of. It acknowledges linguistic knowledge, *longue durée* sedimentations, and routines as fundamental to the temporal organization of spoken language (Couper-Kuhlen and Selting, 2018). By fueling participants’ expectations, sedimented routines enable participants to project what comes next. At the same time, they lay the grounds for improvisation and breach of expectations (Auer and Pfänder, 2011a)—and for a mutual incorporation of linguistic and embodied structures and their potential grammaticalization over historical time.

My observations reverberate with Streeck’s discussion of the parallels between grammaticalization in language and the emancipation of gestures. Streeck observes that “grammaticalization gives us a model how to approach the issue of gesture’s (ongoing) evolution” (Streeck, 2021: 110)—and by extension, it may also give us a model how to approach grammar-body couplings investigated in this paper. Streeck emphasizes parallels in the evolution of gesture and language, but he does not claim that gestures are grammaticalizing. Instead, he suggests that the processes observable in gestures and in spoken languages “are broadly characteristic of human cultural and symbolic evolution” (Streeck, 2021: 01, footnote 1). This leaves open the status of grammar-body-couplings: Are they composed of structures that evolve in parallel, or are they integrated into a whole and undergo change, routinization, social sedimentation, and eventually grammaticalization? A first answer to this question is given in this paper: to distinguish between *ad hoc* assembled, ephemeral grammar-body-gestalts, and socially sedimented multimodal

constructions that have grammaticalized the embodied context of their use over time. While repetition and *joint routinization* of an ephemeral gestalt may lead to the local sedimentation of that gestalt among participants who are mutually engaged in shared activities, *collective routinization* emerges across time and space among social groups whose members are not mutually aware of one another. From here, a practice may or may not start to move along the grammaticalization path (Hopper and Traugott, 2003).

## DATA AVAILABILITY STATEMENT

The video data for this study are not publicly available.

## ETHICS STATEMENT

Ethical review and approval was not required for this study. Informed consent was obtained from all participants.

## AUTHOR CONTRIBUTIONS

AS developed the theoretical framework, collected the data with the support of her team, carried out the empirical analysis, and wrote the article.

## ACKNOWLEDGMENTS

I thank the editors of this special issue for their invitation to contribute to this fascinating topic, and for embarking on the adventure of editing it in this format. I also thank the reviewers for critical reading and helpful comments on a previous version of this paper. I am indebted to the Freiburg Institute for Advanced Studies (FRIAS), University of Freiburg/Breisgau, for support of my research on deixis in face-to-face interaction.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2021.662240/full#supplementary-material>

## REFERENCES

- Arminen, I., Koskela, I., and Palukka, H. (2014). Multimodal Production of Second Pair Parts in Air Traffic Control Training. *J. Pragmatics* 65, 46–62. doi:10.1016/j.pragma.2014.01.004
- Auer, P. (2005). Projection in Interaction and Projection in Grammar. *Text & Talk* 25, 7–36. doi:10.1515/text.2005.25.1.7
- Auer, P. (2006). “Construction Grammar meets Conversation: Einige Überlegungen am Beispiel von ‘so’-Konstruktionen,” in *Konstruktionen in der Interaktion Linguistik*. Editors S. Günthner, and W. Imo (Berlin, New York: de Gruyter), 291–314.
- Auer, P. (2009a). On-line Syntax: Thoughts on the Temporality of Spoken Language. *Lang. Sci.* 31, 1–13. doi:10.1016/j.langsci.2007.10.004
- Auer, P. (2009b). Projection and Minimalistic Syntax in Interaction. *Discourse Process.* 46, 180–205. doi:10.1080/01638530902728934
- Auer, P. (2015). “The Temporality of Language in Interaction: Projection and Latency,” in *Temporality in Interaction*. Editors A. Deppermann and S. Günthner (Amsterdam, Philadelphia: John Benjamins), 27–56. doi:10.1075/slsi.27.01aue
- Auer, P., and Pfänder, S. (2011a). “Constructions: Emergent or Emerging,” in *Constructions: Emerging and Emergent*. Editors P. Auer and S. Pfänder (Berlin, New York: De Gruyter), 1–21.

- Auer, P., and Pfänder, S. (Editors) (2011b). *Constructions: Emerging and Emergent* (Berlin, Boston: De Gruyter).
- Auer, P., and Stukenbrock, A. (2018). When 'You' Means 'I': The German 2Nd Ps.Sg. Pronoun Du between Generativity and Subjectivity. *Open Linguistics* 4, 280–309. doi:10.1515/opli-2018-0015
- Balantani, A. (2021). Reference Construction in Interaction: The Case of Type-Indicative "so". *J. Pragmatics* 181, 241–258. doi:10.1016/j.pragma.2021.05.024
- Barske, T., and Golato, A. (2010). German So: Managing Sequence and Action. *Text & Talk* 30, 245–266. doi:10.1515/text.2010.013
- Bühler, K. (1990[1934]). *Theory of Language*. Amsterdam: John Benjamins.
- Bybee, J. (2014). "Cognitive Processes in Grammaticalization," in *The New Psychology of Language. Cognitive and Functional Approaches to Language Structure*. Editor M. Tomasello (Mahwah, NJ: Lawrence Erlbaum Associates).
- Couper-Kuhlen, E. (2018). Finding a Place for Body Movement in Grammar. *Res. Lang. Soc. Interaction* 51, 22–25. doi:10.1080/08351813.2018.1413888
- Couper-Kuhlen, E., and Selting, M. (2018). *Interactional Linguistics: Studying Language in Social Interaction*. Cambridge: Cambridge University Press.
- De Stefani, E. (2014). "Establishing Joint Orientation towards Commercial Objects in a Self-Service Store: How Practices of Categorisation Matter," in *Interacting With Objects*. Editors M. Nevile, P. Haddington, T. Heinemann, and M. Rauniomaa (Amsterdam: John Benjamins), 271–294. doi:10.1075/z.186.12ste
- De Stefani, E., and Gazin, A.-D. (2014). Instructional Sequences in Driving Lessons: Mobile Participants and the Temporal and Sequential Organization of Actions. *J. Pragmatics* 65, 63–79. doi:10.1016/j.pragma.2013.08.020
- Deppermann, A. (2006). "Deontische Infinitivkonstruktionen: Syntax, Semantik, Pragmatik und interaktionale Verwendung," in *Konstruktionen in der Interaktion*. Editors S. Günthner and W. Imo (Berlin, New York: De Gruyter), 239–262.
- Deppermann, A. (2015). "Retrospection and Understanding in Interaction," in *Temporality in Interaction*. Editors A. Deppermann and S. Günthner (Amsterdam: John Benjamins), 57–94. doi:10.1075/slsi.27.02dep
- Deppermann, A. (2018a). "Changes in Turn-Design over Interactional Histories - the Case of Instructions in Driving School Lessons," in *Time in Embodied Interaction: Synchronicity and Sequentiality of Multimodal Resources*. Editors A. Deppermann and J. Streeck (Amsterdam: John Benjamins), 293–324. doi:10.1075/pbns.293.09dep
- Deppermann, A. (2018b). Editorial: Instructions in Driving Lessons. *Int. J. Appl. Linguist* 28, 221–225. doi:10.1111/ijal.12206
- Deppermann, A. (2018c). Instruction Practices in German Driving Lessons: Differential Uses of Declaratives and Imperatives. *Int. J. Appl. Linguist* 28, 265–282. doi:10.1111/ijal.12198
- Deppermann, A., and Günthner, S. (Editors) (2015). *Temporality in Interaction* (Amsterdam: John Benjamins).
- Deppermann, A., and Streeck, J. (Editors) (2018). *Time in Embodied Interaction* (Amsterdam: John Benjamins).
- Deppermann, A., and Pekarek Doehler, S. (2021). Longitudinal Conversation Analysis - Introduction to the Special Issue. *Res. Lang. Soc. Interaction* 54, 127–141. doi:10.1080/08351813.2021.1899707
- Deppermann, A., and Schmidt, A. (2021). How Shared Meanings and Uses Emerge over an Interactional History: Wabi Sabi in a Series of Theater Rehearsals. *Res. Lang. Soc. Interaction* 54, 203–224. doi:10.1080/08351813.2021.1899714
- Diessel, H. (1999). *Demonstratives. Form, Function, and Grammaticalization*. Amsterdam: John Benjamins.
- Diessel, H. (2006). Demonstratives, Joint Attention, and the Emergence of Grammar. *Cogn. Linguistics* 17, 463–489. doi:10.1515/cog.2006.015
- Diessel, H., and Coventry, K. R. (2020). Demonstratives in Spatial Language and Social Interaction: An Interdisciplinary Review. *Front. Psychol.* 11. doi:10.3389/fpsyg.2020.555265
- Evans, B., and Lindwall, O. (2020). Show Them or Involve Them? Two Organizations of Embodied Instruction. *Res. Lang. Soc. Interaction* 53, 223–246. doi:10.1080/08351813.2020.1741290
- Fillmore, C. (1997). *Lectures on Deixis*. Stanford, CA: University of Chicago Press.
- Ford, C. E., and Fox, B. A. (2015). "Ephemeral Grammar: At the Far End of Emergence," in *Temporality in Interaction*. Editors A. Deppermann and S. Günthner (Amsterdam, Philadelphia: John Benjamins), 95–120. Available at: <https://www.jbe-platform.com/content/books/9789027268990-slsi.27.03for>. doi:10.1075/slsi.27.03for
- Goffman, E. (1971). *Relations in Public: Microstudies of the Public Order*. New York: Basic Books.
- Goffman, E. (1981). *Forms of Talk*. Philadelphia: University of Pennsylvania Press.
- Golato, A. (2000). An innovative German quotative for reporting on embodied actions: Und ich so/und er so 'and i'm like/and he's like'. *J. Pragmatics* 32, 29–54. doi:10.1016/s0378-2166(99)00030-2
- Goodwin, C. (1986). Restarts, Pauses, and the Achievement of a State of Mutual Gaze at Turn-Beginning. *Sociological Inq.* 50, 272–302. doi:10.1111/j.1475-682x.1980.tb00023.x
- Goodwin, C. (1994). Professional Vision. *Am. Anthropologist* 96, 606–633. doi:10.1525/aa.1994.96.3.02a00100
- Goodwin, C. (1996). "Transparent Vision," in *Interaction and Grammar*. Editors E. Ochs, E. A. Schegloff, and S. A. Thompson (Cambridge: Cambridge University Press), 370–404. doi:10.1017/cbo9780511620874.008
- Goodwin, M. H., and Goodwin, C. (2012). Car Talk: Integrating Texts, Bodies, and Changing Landscapes. *Semiotica* 2012, 257–286. doi:10.1515/sem-2012-0063
- Haiman, J. (1994). "Ritualization and the Development of Language," in *Perspectives on Grammaticalization*. Editor W. Pagliuca (Amsterdam, Philadelphia: John Benjamins), 3–28. doi:10.1075/cilt.109.07hai
- Heath, C. (1986). *Body Movement and Speech in Medical Interaction*. Cambridge: Cambridge University Press.
- Hopper, P. (1987). Emergent Grammar. *Ann. Meet. Berkeley Linguist. Soc.* 13, 139–157. doi:10.3765/bls.v13i0.1834
- Hopper, P. J. (2011). "Emergent Grammar and Temporality in Interactional Linguistics," in *Constructions: Emerging and Emergent*. Editors P. Auer and S. Pfänder (Berlin; Boston: De Gruyter), 22–44.
- Hopper, P. J. (2015). "Hermann Paul's Emergent Grammar," in *Hermann Paul's "Principles of Language History" Revisited*. Editors P. Auer, R. W. Murray, and P. J. Hopper (Berlin, Boston: De Gruyter), 237–255. doi:10.1515/9783110348842-012
- Hopper, P. J., and Traugott, E. C. (2003). *Grammaticalization*. 2nd Edn. Cambridge: Cambridge University Press.
- Keevallik, L. (2015). "Coordinating the Temporalities of Talk and Dance," in *Temporality in Interaction*. Editors A. Deppermann and S. Günthner (Amsterdam: John Benjamins), 309–336. doi:10.1075/slsi.27.10kee
- Keevallik, L. (2018a). What Does Embodied Interaction Tell Us about Grammar. *Res. Lang. Soc. Interaction* 51, 1–21. doi:10.1080/08351813.2018.1413887
- Keevallik, L. (2018b). "The Temporal Organization of Conversation while Mucking Out a Sheep Stable," in *Time in Embodied Interaction: Synchronicity and Sequentiality of Multimodal Resources*. Editors A. Deppermann, and J. Streeck (Amsterdam, Philadelphia: John Benjamins), 97–122. doi:10.1075/pbns.293.03kee
- Keisanen, T. (2012). "Uh-oh, We Were Going There": Environmentally Occasioned Noticings of Trouble in In-Car Interaction. *Semiotica* 2012, 197–222. doi:10.1515/sem-2012-0061
- Kendon, A. (2004). *Gesture. Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Laberge, S., and Sankoff, G. (1979). "Anything You Can Do," in *Syntax and Semantics, Vol 12: Discourse and Syntax*. Editor T. Givón (New York: Academic Press), 419–440. doi:10.1163/9789004368897\_018
- Lerner, G. H. (1995). Turn Design and the Organization of Participation in Instructional Activities. *Discourse Process.* 19, 111–131. doi:10.1080/01638539109544907
- Levinson, S. C. (2005). *Pragmatics*. Cambridge: Cambridge University Press.
- Lindwall, O., and Ekström, A. (2012). Instruction-in-Interaction: The Teaching and Learning of a Manual Skill. *Hum. Stud.* 35, 27–49. doi:10.1007/s10746-012-9213-5
- Mondada, L. (2016). Challenges of Multimodality: Language and the Body in Social Interaction. *J. Sociolinguistics* 20, 336–366. doi:10.1111/josl.1\_12177
- Mondada, L. (2014a). "Cooking Instructions and the Shaping of Things in the Kitchen," in *Interacting With Objects: Language, Materiality, and Social Activity*. Editors M. Nevile, T. Heinemann, and M. Rauniomaa (Amsterdam: John Benjamins), 199–226. doi:10.1075/z.186.09mon
- Mondada, L. (2014b). Instructions in the Operating Room: How the Surgeon Directs Their Assistant's Hands. *Discourse Stud.* 16, 131–161. doi:10.1177/1461445613515325



- Mondada, L. (2015). "Multimodal Completions," in *Temporality in Interaction*. Editors A. Deppermann and S. Günthner (Amsterdam: John Benjamins), 267–308. doi:10.1075/slsi.27.09mon
- Ningelgen, J., and Auer, P. (2017). Is There a Multimodal Construction Based on Non-deictic So in German. *Linguistics Vanguard* 3, 1–15. doi:10.1515/lingvan-2016-0051
- Pekarek Doehler, S. (2021). How Grammar Grows Out of Social Interaction: From Multi-Unit to Single Unit Question. *Open Linguistics*.
- Pekarek Doehler, S., and Balaman, U. (2021). The Routinization of Grammar as a Social Action Format: A Longitudinal Study of Video-Mediated Interactions. *Res. Lang. Soc. Interaction* 54, 183–202. doi:10.1080/08351813.2021.1899710
- Rauniomaa, M., Lehtonen, E., and Summala, H. (2018). Noticing with Instructional Implications in Post-Licence Driver Training. *Int. J. Appl. Linguist* 28, 326–346. doi:10.1111/ijal.12199
- Schütz, A., and Luckmann, T. (1973). *The Structures of the Life-World*. London: Heinemann.
- Stivers, T., and Rossano, F. (2010). Mobilizing Response. *Res. Lang. Soc. Interaction* 43, 3–31. doi:10.1080/08351810903471258
- Selting, M., and Couper-Kuhlen, E. (Editors) (2001). *Studies in Interactional Linguistics* (Amsterdam; Philadelphia: Benjamins).
- Streeck, J. (1988). The Significance of Gesture. *IPrAPiP* 2, 60–83. doi:10.1075/iprapip.2.1-2.03str
- Streeck, J. (1995). "On Projection," in *Social Intelligence and interaction. Expressions and Implications of the Social Bias in Human Intelligence*. Editor E. N. Goody (Cambridge: Cambridge University Press), 87–110. doi:10.1017/cbo9780511621710.007
- Streeck, J. (2002). Grammars, Words, and Embodied Meanings: On the Uses and Evolution of So and like. *J. Commun.* 52, 581–596. doi:10.1111/j.1460-2466.2002.tb02563.x
- Streeck, J. (2007). "Geste und verstreichende Zeit," in *Gespräch als Prozess. Linguistische Aspekte der Zeitlichkeit verbaler Interaktion Studien zur deutschen Sprache*. Editor H. Hausendorf (Tübingen: Narr), 157–180.
- Streeck, J. (2009). Forward-Gesturing. *Discourse Process*. 46, 161–179. doi:10.1080/01638530902728793
- Streeck, J. (2018). Grammaticalization and Bodily Action: Do They Go Together. *Res. Lang. Soc. Interaction* 51, 26–32. doi:10.1080/08351813.2018.1413889
- Streeck, J. (2021). The Emancipation of Gestures. *Il* 1 (1), 90–122. doi:10.1075/il.20013.str
- Streeck, J., Goodwin, C., and LeBaron, C. (2011). *Embodied Interaction: Language and Body in the Material World*. Cambridge: Cambridge University Press.
- Stivers, T., and Sidnell, J. (Editors) (2012). *Handbook of Conversation Analysis* (Malden, MA: Wiley-Blackwell).
- Stukenbrock, A. (2008). "Wo ist der Hauptschmerz?" - Zeigen am eigenen Körper in der medizinischen Kommunikation. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 9, 1–33.
- Stukenbrock, A. (2010). Überlegungen zu einem multimodalen Verständnis der gesprochenen Sprache am Beispiel deiktischer Verwendungsweisen des Ausdrucks "so. *InLiSt - Interaction and Linguistic Structures* 47, 1–23.
- Stukenbrock, A. (2014). Take the Words Out of My Mouth: Verbal Instructions as Embodied Practices. *J. Pragmatics* 65, 80–102. doi:10.1016/j.pragma.2013.08.017
- Stukenbrock, A. (2015). *Deixis in der Face-to-Face-Interaktion*. Berlin, Boston: De Gruyter.
- Stukenbrock, A. (2017). "Intercorporeal Phantasms: Kinesthetic Alignment with Imagined Bodies in Self-Defense Training," in *Intercorporeality: Emerging Socialities in Interaction*. Editors C. Meyer, J. Streeck, and J. S. Jordan (Oxford: Oxford University Press), 237–263.
- Stukenbrock, A. (2018a). "Forward-Looking: Where Do We Go With Multimodal Projections?," in *Modalities and Temporalities: Convergences and Divergences of Bodily Resources in Interaction*. Editors A. Deppermann and J. Streeck (Amsterdam: John Benjamins), 31–68. doi:10.1075/pbns.293.01stu
- Stukenbrock, A. (2018b). "Mobile Dual Eye-Tracking in Face-to-Face Interaction: The Case of Deixis and Joint Attention," in *Eye-Tracking in Interaction*. Editors G. Bröne and B. Oben (Amsterdam: John Benjamins), 265–302. doi:10.1075/ais.10.11stu
- Stukenbrock, A. (2020a). Deixis, Meta-Perceptive Gaze Practices, and the Interactional Achievement of Joint Attention. *Front. Psychol.* 11. doi:10.3389/fpsyg.2020.01779
- Stukenbrock, A. (2020b). "Mit Blick auf die Geste - Multimodale Verfestigungen in der Interaktion," in *Verfestigungen in der Interaktion*. Editors B. Weidner, K. König, W. Imo, and L. Wegner (Berlin, Boston: De Gruyter), 231–262. doi:10.1515/9783110637502-010
- Stukenbrock, A., and Dao, A. N. (2019). "Joint Attention in Passing: What Dual Mobile Eye Tracking Reveals about Gaze in Coordinating Embodied Activities at a Market," in *Embodied Activities in Face-To-Face and Mediated Settings*. Editors E. Reber and C. Gerhardt (Basingstoke: Palgrave Macmillan), 177–213. doi:10.1007/978-3-319-97325-8\_6
- Svensson, M. S., Luff, P., and Heath, C. (2009). Embedding Instruction in Practice: Contingency and Collaboration during Surgical Training. *Sociol. Health Illness* 31, 889–906. doi:10.1111/j.1467-9566.2009.01195.x

**Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Stukenbrock. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.