What Does Embodied Interaction Tell Us About Grammar?

Leelo Keevallik

This is an electronic version of an article published in:
Keevallik, L., (2018), What Does Embodied Interaction Tell Us About Grammar?, *Research on Language and Social Interaction*, 51(1), . https://doi.org/10.1080/08351813.2018.1413887

What does embodied interaction tell us about grammar?

(The paper has benefited from numerous kindly critical remarks by Charles Antaki, Mathias Broth, Elise Kärkkäinen, Per Linell, Nigel Musk, Charlotta Plejert, Sandra Thompson)

Leelo Keevallik

Abstract
This paper navigates the findings of conversation analysis, interactional linguistics, and related multimodal studies in order to summarize what we know about the grammar-body interface. It shows how grammar is fitted to sequences and trajectories of embodied activities, as well as deployed interchangeably with bodily displays, resulting in truly multimodal patterns that emerge in real time. These findings problematize both the paradigmatic and syntagmatic structures documented in verbal-only linguistics. They call for a reconceptualization of grammar as an assembly of routinized methods for the organization of vocal conduct, capable of incorporating aspects of participants' bodily behavior.

1. Grammar for multimodal interaction

The term *grammar* refers to the composition of clauses, phrases and words in a natural language. There is, however, a serious disagreement in the field of linguistics between scholars who see patterns of grammar as stemming from individual cognition, and those who argue that they are epiphenomena based on speakers' interactional experience. From a formal linguistic perspective, grammar has been thought of as a device for organizing information in self-contained sentences that coherently express propositions. This has resulted in the description of such useful structures as phrases, predicates, argument structure, and clauses, but as if they were separate from actual usage. It has been considered sufficient to work with written language, and to contemplate the "mathematic" esthetic of combinatory sentences, such as "Either if the girl eats ice cream then the boy eats ice cream, or if the girl eats ice cream then the boy eats candy", in order to construct a decontextualized grammar (Pinker 1995: 92–97). The alternative, functional linguistics, has instead documented motivations for lexical and syntactic features from outside the "language system" itself (Bybee 2003) and argued that social and interpersonal matters are central for understanding the nature of language. An utterance that seems ungrammatical when imagined, such as "if the guys do [pause] then the ladies will do [pause]" makes perfect sense in its actual context where it is timed with embodied demonstrations, as will be argued below (example 10). Accordingly, functional linguistics has embraced audio and video recordings of people interacting in various situations.

Among other things, a functional approach calls us to re-consider the central linguistic categories, mostly established on the basis of written language, such as noun phrases and (complex) clauses, which may be even hard to locate in interaction. From a functional perspective, these categories should furthermore be complemented with interactional linguistic structures, such as those used for repair initiation (see a recent attempt by Ginzburg & Poesio 2016), increments, turn-transition, and responding. These

devices also constitute grammar, as they form conventionalized patterns that can be abstracted and systematically described together with their social function. Similar to the rest of grammar and lexis they must be learned within every single language community. At the same time, they are characteristically sensitive to temporal contingencies and interpersonal concerns. In numerous spoken as well as written genres we produce grammar in real time. Therefore, the precise emergence of linguistic patterns (Hopper 1998) should be of central theoretical interest to linguists to at least counterbalance the formal and often excessively hierarchical models of recalled or imagined atemporal sentences. Furthermore, language is but one resource of sense-making. As interacting human beings we cannot merely rely on our earlier experiences of lexis and grammar, because this abstracted knowledge does not in itself guarantee mutual understanding here and now. A more realistic view of the achievement of intersubjectivity is to be found in the complex interplay between language, body movements, and the material environment (Linell 2009, Streeck et al. 2011). Thus, both social action formation and the emergence of grammatical structure are multimodal processes.

One branch of functionalism that has taken a determined stance against formalist reductionism is interactional linguistics, established as a rigorously empirical research endeavor around the turn of the millennium (Ochs et al. 1996, Selting & Couper-Kuhlen 2001, Couper-Kuhlen & Selting, to appear). Methodologically it relies on conversation analysis, pedagogically summarized in Sidnell (2010) and Clift (2016). It embraces the basic conversation analytic findings on the turn-taking system and, crucially for linguistics, the nature of temporally emerging and negotiable turn-constructional units (Sacks et al. 1974). Analytically it profits from the findings on the sequential organization of conversational actions (Schegloff 2007), which enables the study of the systematicity between syntactic form and its function as oriented to by participants. This paper navigates the findings of interactional linguistics and related multimodal studies, collecting some fruits of the recent "embodied turn in research on language and social interaction" (Nevile 2015). It targets the emergence of grammar as affected by body movements, both locally and in the form of routinized structures, asking how embodied behavior should figure in our description of grammar. The paper will follow the basic logic of conversational interaction, from sequences and actions to the formatting of turn-constructional units. First, embodied actions construct sequences (Section 2), thus shaping the ensuing turns-at-talk. Second, embodied conduct is also essential in the functioning of non-sequence-organized turns (Section 2.4). Third, embodied demonstrations are used for constructing turns, and turn-constructional units (Section 3), both fundamental structures of interaction. Finally, it will be shown how grammatical structures emerge within multimodal streams of interaction (Section 4).

The notion of embodiment is here limited to body movements (see Nevile (2015) on the large variety of aspects often lumped together as 'multimodality', and Gallagher (2011) on a similar overview on 'embodiment'). Since the focus will be on the interplay of grammar and the body, often as a whole, the embodied nature of language production itself will mostly remain untouched. Studies of interactional prosody (starting from Couper-Kuhlen & Selting 1996) have already formed an influential stream of research, with a focus on action formation and turn-construction. Likewise, studies on the co-production of talk and gesture is a powerful branch by itself (e.g. Kendon 2004, Streeck 2009), in particular targeting deixis and reference (Mondada 2009a,b, Eriksson 2009).

Gesture also holds a distinct position among the different interactional resources, as it is tightly connected to language production (Schegloff 1984, McNeill 2005). This research cannot be adequately surveyed in the current paper, which is mostly concerned with embodiment beyond gaze and gesture. Space, objects, and the material surroundings will not be in focus either, but they will of course figure in the different analyses. *The objective here is to summarize what we know about the relationship between grammar and its categories on the one hand, and interactionally relevant bodily actions on the other, asking why grammarians should care about the speakers' bodies beyond their brains and vocal tracts*.

2. Embodied sequential action and grammar

There are two central tenets in analyzing grammar in interaction: (1) language is an intrinsically temporal phenomenon that emerges in turns-at-talk, and (2) human action is not exclusively carried out in the verbal domain.

In the seminal paper "Turn organization" Emanuel Schegloff formulated the concept of *positionally sensitive grammars* (1996: 108–111), embracing the temporal nature of linguistic patterns. He argued that instead of starting a grammatical analysis from intentions and propositions, one should analyze the actual moment when a speaker begins an utterance, because its grammatical composition is often occasioned by the immediately preceding talk. He suggested exploring "the sorts of [conversational] junctures and contingencies which the organizations of interaction engender on the one hand, and the forms of grammatical structure and practice which get deployed at those junctures and in those contingencies on the other" (Schegloff 1996: 111). In short, grammatical structure is a result of sequential contingencies and interactional function.

In addition, social interaction is inherently multimodal (see special issues edited by Stivers & Sidnell (2005), Deppermann (2013a), Rasmussen et al. (2014)). People use grammar, the body, space, and material objects in various combinations to build coherent courses of action (Goodwin 2000). Several studies have considered the parallel deployment of grammar and the body to achieve e.g. (mutual attention for) assessments (Goodwin & Goodwin 1987, 1992, Lindström & Mondada 2009), reference in offers of concrete objects (Kärkkäinen & Keisanen 2012), and compliments (Keisanen & Kärkkäinen 2014). Furthermore, Ruusuvuori and Peräkylä (2009) showed the relevance of timing: facial expressions can transgress verbal assessments, providing extended opportunities for affiliation, and stretching the temporal boundaries of an action. These studies provide important insights into multimodal action formation. In the current section we will focus on sequences where different participants implement actions in different modalities, which affects the grammatical resources deployed.

2.1. Embodied responses and grammar

Sequentially relevant action can be accomplished without language, which is easy to see in some types of responsive actions. A request can, and often has to, be granted in an embodied manner. In co-present settings requests for concrete objects are typically granted without verbal acceptance (Rauniomaa & Keisanen 2012, Mondada 2014a), responses to instructions in air traffic control training are necessarily embodied (Arminen et al. 2014). Responses to compliments can also be embodied (Keisanen &

Kärkkäinen 2014). Interestingly, embodied and verbal responses may diverge in their functions. As Stivers (2008) has shown for English, the recipient's nodding in mid-telling indicates access to the teller's stance, while vocal continuers simply align with the activity in progress. Accordingly, they are positioned and treated differently, with nodding used when the story-teller has revealed her stance and allowing her to move to the next component of the telling. Along similar lines, bearing in mind Schegloff's (1996) idea of positionally sensitive grammars, any embodied sequential action is consequential for the formatting of the next action. Thus, an utterance after an embodied response is fitted to it, and we should explore how.

2.2. Grammar after an embodied response

Among other things, when a conditionally relevant embodied response to a first action has been produced, the next turn can still be formatted as a 'sequence-closing third', as described for verbal-only sequences (Schegloff 2007: 118, 123–127). A recent example of these sequence-closing thirds in French can be found in Mondada (2014b: 142). She discusses sequences in a surgical team consisting of an instruction, an embodied instructed action, and the affirmative particle *ouais* 'yeah'. As shown in example (1) the surgeon produces the instruction to take a piece of tissue (line 1), the assistant accomplishes the required action with his pliers, marked with asterisks. The surgeon ratifies it with a *ouais* (line 4).

(1) An acknowledging 'third' after an embodied response: Surgeon and an assistant (adapted from Mondada 2014b: 142)

```
1 Surgeon:            .h reprend plus *près lá::*
                      'Takes closer there again'
2 Assistant                           *((drops previ))*
3                     *(0.2)
  Assistant           *((takes the tissue))*
4 Surgeon:             ouais
                       'Yeah'
6                     (0.5)*
```

This kind of instruction sequence frequently occurs in teaching activities, as classically described in classroom interaction: initiation by teacher – response by students – feedback by teacher (Sinclair & Coulthard 1975, Walsh 2011: 17–20). When the student response is embodied, the teacher's feedback is reactive to, and timed with that and not with any prior spoken utterance. An example from a context where embodied responses abound, a dance class, follows (example 2, in Swedish). The instruction has been for each dance couple to enter the dance floor on the beat after another couple has stopped. Every time a couple succeeds, the two teachers A and B shout out feedback (lines 2,3,5,7,8).

(2) Sequence-closing thirds after embodied responses: Dance teachers (author's corpus)

```
1          ((couple 1 stops, couple 2 enters))
2 TeaA:    JEEe.=
           'Yeah'
3 TeaB:    =BRA, just de.
           'Good. that's it.'
4          ((couple 3 stops, couple 4 enters))
5 TeaB:    JUSt de.
           'That's it.'
```

```
6          ((couple 5 stops, couple 6 enters))
```

Fig. 1: Couple 6 enters from the left, teacher B standing in the middle, teacher A behind the leaving couple 5.


```
7 TeaB:    #1 B[RAa.]
               'Good.'
8 TeaA:        [JEE]ee.
               'Yeah'
```

The evaluative *jeee* 'yeah', *bra* 'good', and *just de* 'that's it' are temporally fitted with the successful students' performance. The range of feedback formats depends on the activity and the degree of appreciation conveyed. It is nevertheless limited to certain structures, such as adjective and adverbial phrases and special fixed formats involving deixis (such as *just det*, lit. 'exactly that' (line 5), *bra där* 'good there', and *där satt det* 'that's it', lit. 'there it was'). Similarly short and structurally limited items, *voilà* 'that's it' and *perfetto* 'perfect' emerge in Swiss Italian data from driving classes, as responses to technology-mediated embodied action (De Stefani and Gazin 2014; on camera-mediated embodied responses at a French TV studio, see Broth 2014). All these evaluative structures treat the prior embodiment as a sequentially relevant action, and cannot be analyzed without it. They are structurally formatted as a special kind of a sequence-closing third, so-called *musings* that "offer either some analysis/diagnosis of the prior sequence or some assessment of it" (Schegloff 2007: 142–148). It is highly likely that the syntactic and lexical systematicity of musings enables the kind of paradigmatic analysis that has always been inherent to linguistics, while respecting the participants' local orientation to different kinds of embodied actions. This will modify, among other things, what we mean by grammar, since qualitatively different, perhaps superficially less complex, structures will be brought to attention, structures that have simply been ignored as irregular, formulaic, or incomplete. Ironically, this illustrates how grammar in many of its conceptualizations is far too narrow to describe language structures and the interactional regularities that govern them, because it is too focused on the abstractions of imagined completeness.

2.3. Embodied initiating actions and following turns

Similar to embodied second actions, responses to embodied initiating actions are formulated in a positionally sensitive way and feature specific methods for accomplishing social action. Accordingly, a grammatical analysis should take the embodied initiating action, and not an abstract proposition as its starting point. Entirely embodied initiating actions have so far been identified in repair initiations, directives, and offers.

Seo and Koshik (2010) describe how two body postures, the "head poke" and the head tilt, can be used to initiate repair in a classroom. Most recently, Mortensen (2016) has added the cupping of a hand behind the ear to this list, showing how it functions as an initiating turn at talk when produced at a transition relevance place in second language classrooms. The ensuing vocal action accomplishing the repair is in these cases grammatically fitted to the embodied initiating action, and is structurally limited in ways similar to responses to verbal repair initiations expressing hearing trouble. That is, they feature repetitions and reformulations.

Another initiating action that has recently drawn attention is requests and, more broadly, recruitments. Recruitments have been conceptualized as ways in which assistance may be sought or in which we perceive another's need and volunteer assistance (Kendrick & Drew 2016). Crucially, Kendrick and Drew (2016: 8–9) demonstrate how an embodied display of trouble may result in embodied assistance, thus not necessitating language at all. Along similar lines, Rossi (2014) argues that requests can be accomplished without language when the requested action is projectable from the advancement of an activity. Among his Italian cases, we can find an instance when an embodied first action occasions an inquiry (example 3): Lucio apparently does not understand what is requested from him by embodiment only, described in line (3). His following turn, a clarification question, is grammatically sensitive to the fact that an initiating action, some kind of a request, has already taken place: it is a first person modal interrogative. He complies with the requested action only after the clarification sequence has come to a close (in line 8), so the adjacency pair in lines (5,6) can be seen in parallel to a verbal insertion sequence. Here it is "inserted" into an embodied recruitment sequence, featured in lines (3, 8).

(3) A clarification sequence after an embodied initiating action: Filling out forms (adapted from Rossi 2014: 321-322).

```
1          (4.7)
2 Flora    ((adds her signed forms to the pile))
3          ((places blank form and pen on the table next to Lucio))
4          (1.0)

5 Lucio:   cosa devo     fare
           what must-1SG do-INF
           'What should I do?'

6 Flora:   eh  anche tu      devi     scrivere  il  tuo  nome
           PAR also  2SG-NOM must:2SG write-INF the your name

           firmare  e   la  data
           sign-INF and the date
           'Well you too must write your name, sing, and (put) the date'
7          ((pushes form closer to Lucio))
8 Lucio    ((grabs form and pen))
```

In a similar Italian everyday setting, Rossi (2012) has targeted the function of two specific grammatical structures, imperatives and interrogatives with *mi* 'to/for me'. He demonstrates that imperatives are selected to implement requests that are integral to the already established joint project. The project can be embodied, such as blowing one's nose. In example (4) Dino (female) suggests that Sergio (male) blow his nose, so within this project initiated by her it is legitimate for Sergio to address Dino with the imperative *tieni* 'hold' in order to free his hand of the dye bottle.

(4) Grammatical form fitted to embodied activity trajectory (adapted from Rossi 2012: 437)

```
1 Dino     ((raises paper to Sergio's nose))
2 Sergio   ((brings free hand to nose))
3          ((positions dye bottle in a way suitable for Dino to grab it))

4 Sergio:  tieni          questo,
           hold-NPST-2SG this
           'Hold this,'

5 Dino     ((grabs the bottle))
6 Sergio   ((blows nose))
```

We are thus currently witnessing an exciting opening towards understanding the choice between alternative grammatical formats as based on the ongoing embodied activities. Instead of departing from the idea that Sergio wants something to be done and finds a combination of linguistic elements to express that proposition, the analysis should recognize that the format is motivated by the joint trajectory of social action at the very moment when he starts to talk. His bodily problem, his holding of the bottle and Dino's free hand all conspire to make imperative + deictic the appropriate structure to be used. The imperative formulates the action of holding as yet-to-be-done by the recipient, who is targeted as the single doer of the action by this very imperative. The deictic + positioning of the bottle achieves reference to the object in the speaker's hand. It is thus not sufficient to see grammar, e.g. a predicate followed by an object, as an integrated system of combinatory rules, but as a method of solving local problems. *Abstract grammatical categories, such as an imperative or a deictic, arise from the physical co-presence of the participants, the specific arrangement of their bodies (as available or not), and the organization of material world around them*. The actual combination of an imperative and a deictic in the verbal stream is almost coincidental to the embodied trajectory of handing over the bottle to an available other, while of course the actual temporal ordering of a predicate and an object has been conventionalized differently in different languages, reflecting the routine nature of such recruitments.

Recently, studies on various service institutions have revealed grammatical regularities of requests that are contingent on the physical distance between the participants or they co-occur with the manipulation of an object. In Finnish kiosk encounters, requests for tobacco products take the form of a noun phrase, unless the customer is still approaching the counter, in which case a full clause is used (Sorjonen & Raevaara 2014). The study contrasts phrases such as *Pieni punainen ällämmä*. 'Small red LM.' (p. 248) with clauses such as *Anna pieni punainen ällämmä*. 'Give (me) a small red LM' (p. 254). At an American shoe-repair shop, requests formulated as inquiries about the

repairability of shoes are accompanied with minimal manipulation of the object, while requests formulated as solutions are accompanied by manipulations of the shoes to visibly reveal a solution (Fox and Heinemann 2015). *Grammatical choices are thus systematically made in terms of the physical context where they are deployed; they are fitted to features of space as well as to ways of sense-making with focal objects*.

Directives beyond requests have attracted interest in terms of the choice between embodied and verbal devices, primarily in family interaction (Tulbert & Goodwin 2011). For example, Cekaite (2010) has shown that when verbal directives prove to be unsuccessful, parents can resort to physical "shepherding" as a means to orient the child towards the required routine task.

As it happens, almost no analysis so far has targeted verbal responses to embodied initiations, the only exception being Kärkkäinen and Keisanen (2012: 602–607), who show that embodied offers are optionally received with a "thank you". One reason for this research gap may be the more or less fixed formats used in this kind of responsive actions. For a grammatical analysis they may seem as irrelevant – but precisely for this reason, we need a different grammatical theory that fully embraces the more routinized "chunks" of language alongside the other, productive patterns. In fact, the share of (semi-)fixed formats may be larger than has been acknowledged, especially in turn-beginnings, where projective elements have conventionalized formats, and at turn-completion, where stance-taking devices likewise feature numerous formulaic phrases. Crucially for the domain of embodied interaction, if a formulaic phrase regularly combines with a specific bodily conduct, its timing and linguistic details lend themselves to a systematic description of recurrent bodily-vocal constructions. There could be patterns such as a 'thank you' coordinated with the manual accepting of an offered object, a responsive 'hello' coordinated with a hug, or an 'I dunno' uttered during a shrug, as well as many less readily imaginable combinations. A shrug can also be an alternative to the verbal response, as can a head shake for a 'no' in many Western cultures. In short, *responses are either embodied, verbal, or bodily-vocal, which is essential for a systematic description of how language functions, even though here we are likely to discover its more formulaic aspects*.

2.4. Embodiment and verbal actions that are not sequence-organized

While the analysis of social action has heavily relied on adjacency pairs and their expansions, there are "plenty of utterances that are neither first pair parts nor second pair parts" in everyday life (Linell 2009: 185). These utterances produced in co-presence may well turn out to be the best locus for demonstrating how embodiment matters for grammar. It is here we can see how chunks of language that are sensitive to situational relevancies emerge as something we would call a grammatical structure. Grammar is not necessarily organized in terms of sequence, as defined by Schegloff (2007: 1–3). In fact, he makes a difference between the broader sequential organization, and the more narrow conversational sequence organization. In the current section we will be looking at sequentially organized actions in the broader sense.

In order to explore context-sensitivity beyond conversation, an obvious setting seems to be instructional activities that target the body. That is, where one participant mostly uses vocal resources in order to instruct the others' bodies. In studies of different instructional activities, such as dance classes and driving schools, we have already been

able to discover grammatical regularities. For example, Estonian 'no'-prefaced turns formulate students' mistakes in dance classes (Keevallik 2012). Another study of a Swiss Italian driving school has demonstrated the dependence of grammatical formats on the timing of the instruction: the "early" ones involved verbs while the "late" ones did not (DeStefani & Gazin 2014). *These corrective verbal moves both react to, and project, the other's embodied behavior, which again undermines the formal perspective on grammar as the combining of elements to express a proposition*. Instead, grammatical elements such as particles, verbs, and noun phrases are chosen in relation to the temporal structure of the ongoing activity. Characteristically, teacher instruction can be simultaneous with the emerging student responses, as has been shown for directives in Swedish (Broth and Keevallik 2014). Here is an example from an Estonian pilates class. The teacher is standing and prompting the students on the floor. At least some of the students already seem to know the exercise - from a forward bend, they are now supposed to straighten up and then start bending again.

(5) Responsive-performative grammar: Pilates class (author's corpus)

Fig. 2. Fig. 3. Fig. 4.

```
1   j#2:a-          -a::::#3  ülesse          sIR#4::::::gemaks.
    'and'                     'upwards'       'become more straight.'
    ((teacher moving towards the student in the right corner,
    stops in #3))
```

Fig. 5. Fig. 6.

```
2   J↑A#5A:::::.                .hh >ja jäll#6e.<
    'yeah.'                         'and again.'
    ((upward moving gesture         ((teacher returning
    from the student's head))       to the center))
```

The teacher utters a grammatically complex structure *ja ülesse sirgemaks. jaa. ja jälle*. 'and upwards more straight. Yes! and again.'. These words are well-timed with the students' body movements. They instruct and provide rhythm for the exercise: as can be seen in the difference between #1 and #2, the students respond to the teacher's 'and' by starting moving their torsos upwards, so that the descriptive 'upwards' emerges partly as a responsive approval of their bodily action, partly as a simultaneous description, and

partly as a reminder prompt for the students. In contrast, the morphologically complex *sirgemaks* 'become more straight' provides an instruction and critique of the simultaneous action, as the students have straightened up by its first syllable (#3).

Studying surgery, Mondada (2014a) discusses similar request-type actions as if "exploiting" the trajectory of the activity and its projectability, occasionally temporally "merging" with it. The grammar of these non-sequence-bound utterances can reveal systematicity between embodied and verbal behavior at every moment in time. For example, the Estonian morphological endings illative *-sse* in *üle-sse* 'up-to' and translative *–ks* in *sirge-ma-ks* 'straight-more-into' (line 1) are used here in the service of the ongoing concrete movement (guiding its direction and quality), which is typically not in focus in the abstract discussions of mapping meanings to case endings. Boundaries in the pilates exercises are regularly prompted with the conjunction *ja* 'and' (as in *ja ülesse* 'and upwards', *ja jälle* 'and again'). Embodied responsiveness may furthermore provide us with evidence of how participants understand the various structures, so we as analysts do not have to just assume that we all share the same knowledge about them. Crucially, the grammatical combination of a conjunction, the illative and the translative form are here a local result of the students' bodies moving and being instructed in consecutive exercise segments. Linguistic structure, even though thoroughly "grammatical", emerges here not as an application of some cognitive process but in the interplay with others' exercising bodies and the routine organization of the class.

One under-studied domain of reactivity to contextual matters is response cries (Goffman 1981: 78–123). Response cries can attend to embodied experiences by either self or other, featuring both responsiveness and performativity (Keevallik 2016). The prosodically extreme *J↑AA::::* 'yeah' in example (5, line 2) above responds to the students' straightening spines and at the same time encourages them to extend and enhance their efforts, providing a time-frame for the appropriate performance – until the end of the *J↑AA::::* 'yeah'. As can be seen in #5, the first student (far to the left) starts bending down in the middle of the teacher's quick and low *ja jälle*. 'and again'. The *J↑AA::::* has thus accomplished a lengthy and specific response in the students' bodies while being produced, featuring simultaneity of verbal and bodily behavior, distributed across different participants. This is an example of how we can begin to discover the meaning and function of at least affirmative particles, hitherto believed to be primarily confirmations after polar questions, and agreements. Studies on other response cries (*ouch, woho, yuck*, etc.) in co-presence are likely to render similarly complex results (see Keisanen (2012) on *uh-oh*), and *problematize the way linguistics delimits what does and does not belong to lexis, as well as what kinds of meaning can be tied to a lexical item*. Furthermore, it is quite likely that the origin and home-base of various kinds of particles is in embodied and not in verbal-only activities.

2.5. Summary

The above overview has shown that grammar in various ways is shaped by bodily conduct. Syntactic structures usefully described by linguists (interrogative vs. imperative, clause vs. phrase) can be occasioned by an (other's) embodied action; emerging lexis and fixed formats are fitted to local activity trajectories. Likewise, the temporality and spatiality of the embodied action may be consequential for the choice of grammar. *Especially as a counterbalance to the idea that language structure results from*

abstract cognitive principles, we should entertain the possibility that it emerges stepwise, incrementally, as a series of local solutions to participants' current actions, such as practicing a pilates exercise, or learning to drive a car. Patterns can be established retrospectively and, if recurrent, generalized as grammatical constructions that are not necessarily abstract.

Embodied responses to verbal initiating actions are relevant in many co-present settings, such as those featured above (surgery, family interaction, dance classes) and some of them are mediated by technology, such as flight control training (Arminen et al. 2014) or collaborative writing with the help of a spellchecker (Musk 2016). Occasionally, verbal actions can be alternatives to embodied ones, albeit at different moments in the evolving trajectories (Cekaite 2010, Rossi 2014) or in qualitatively limited activities, such as those involving physical objects (Stevanovic & Monzoni 2016). The above branch of research certainly informs us of the general usefulness of grammar as action, but also suggests how regular sets of grammatical formats emerge in real-time embodied interaction, eventually constituting a systematic whole. Furthermore, speakers and analysts alike are capable of abstracting verbal-bodily patterns of sequential action. *The overall result could partially be reminiscent of traditional grammatical paradigms but these paradigms would take action-in-context, not the linguistic system, as their basis.*

It has also become clear that the mutual temporalities of verbal and embodied trajectories are complex (Mondada 2014a,b; Keevallik 2016), and do not necessarily become organized in what could be seen as traditional conversational sequences. Overall, we lack a systematic understanding of grammatical structures that are used reactively and performatively in relation to various embodied contingencies. These structures range from more or less fixed formats to relatively flexible ones, from those that can be organized into paradigms to those that are more or less idiosyncratic outcomes of entirely local contingencies, but nevertheless witness to the adaptability of grammar to the (embodied) task at hand.

3. Embodiment and turn-construction

Moving on from sequence analysis, this section of the paper targets the role of embodiment in the construction of turns. Turns are conversational units that speakers demonstrably orient to and that are regularly built of linguistic materials. A subsection of these materials can be described as clauses or phrases, and it seems that the speakers of some languages at least also orient to these. However, a grammarian interested in describing structures in conversation cannot afford to ignore elements and behavior that go beyond what we have grown accustomed to seeing as a full clause, phrase, or even a word. For example, the category of "interjections" enjoys a strong and extremely varied presence in interaction, occasionally transgressing conventionalized forms but they are nevertheless treated as communicatively meaningful. Crucially, when trying to figure out the order of elements in interaction, including the order of linguistic items in speech, the turn is an empirically grounded unit, while some written-language biased structures, such as a sentence, can be almost impossible to find. A turn is also an essentially temporal structure, which is to be described accordingly in terms of turn-beginning, progression, and completion.

Recent work on interaction has discussed how participants visibly prepare themselves before starting to talk. Turn-design, and especially turn-beginning, arises from local contingencies, "embodied" ones among them. As concisely expressed by Depperman (2013b: 93): "Turn-construction is both responsive to and uses spatial, kinesic [sic], tactile, visual and auditory affordances and constraints of the participants' bodies." Establishing mutual attention is essential in pre-beginnings (Mondada 2009a, Keisanen & Rauniomaa 2012, Mortensen 2009). According to Schegloff (1996: 92–93), interactionally relevant behavior in pre-beginnings includes turning the head towards a potential recipient, the onset of gesture, lip parting, coughing, and a hearable inbreath. Mondada (2007) has shown how pointing gestures during a co-participant's ongoing talk can function as a claim to the next turn, which underlines the centrality of multimodal behavior in turn-taking. However, there has been no systematic study of the grammar used in turn-beginnings with and without having established attention; we only know that restarts may be deployed to attract recipients' gaze (Goodwin 1980). It is thus yet to be discovered how the precise verbal structure of turn-constructional units (TCUs) is fitted to various attention frameworks and pre-beginning behaviors. It is likely to affect the use of turn-initial pragmatic particles and perhaps also the use of grammatical structures that do or do not build on previous syntax.

During a turn, speakers can produce talk simultaneously with embodied displays, including gestures, but they may also perform embodied displays without simultaneous talk. Some of the displays are projected by specific phrases, such as *und ich so/und er so* 'and I'm like/and he's like' in German (Golato 2000) and *it's like* in English (Streeck 2002, Fox & Robles 2010), which suggests that these could be systematically described in a comprehensive grammar. A turn-constructional unit can contain both a projective grammatical phrase and an ensuing embodied display, a depiction (to use a recent term by Clark 2016), as its completion, as shown in example (6). Here, the German phrase in line (2), *ich war echt so* 'I was really like' is followed by an embodied display (R spreads hands and freezes), which is characteristically accompanied by a non-lexical vocalization, in the current case *.ähhhhh*.

(6) Phrase projecting an embodied display (adapted from Golato 2000: 44-45)

```
1 R:       das hat mir tobias alles gesaht also
           'Tobias told me all of that well'

2          Ich war e*cht so .ähh[hhh*
           'I was really like .ahhhhh'
                    *((spreads hands,
                      freezes in position))*
3 T:                           [ich schlief ich
                                'I slept I'
4          sach fahr rechts ran, ganz ruhich
           'said pull over to the right, stay calm'
```

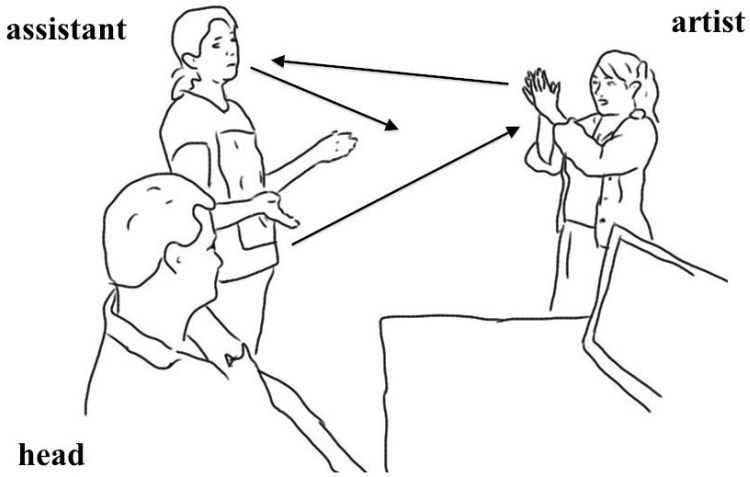
Furthermore, embodied displays may constitute separate units within ongoing turns (Keevallik 2014). Example (7) comes from a meeting at an Estonian theater workshop, and shows a turn where the artist explains her idea about twirling pieces of cloth. In lines 1–5 she argues that it is important to make the cloth's structures appear and

disappear through rotation. In the middle of this complex turn (in line 4) she produces a demonstration that involves a non-lexical vocalization and a gesture.

(7) Embodied demonstration as a TCU, theater workshop (Keevallik 2014)

```
1 Artist:  A   see on     just      SEE lahe et,
           but it  be-3SG precisely it  cool that
           ‘But the thing that is cool is that’
```

Fig. 7: “SEE ‘this’”




```
2          ta k- nagu TEkib      vata. (.)
           it    like appear-3SG PAR/look:2SG
           ‘it appears, y’see’

3 Head:     °[a],°
            ‘uh’

4 Artist:    [d]rr↑rrrr,

5          ja  sis  tõmbub TAgasi nii.
           and then retract-3SG   like.this
           ‘and then retracts like this.’
```

Fig. 8: Pitch contour on “drrrrrr”[1].

[1] There may be a slight reflection of the head’s *a* at the beginning of the pitch curve but his voice is very low in relation to the artist’s voice, which suggests that the contour is nevertheless reliable.

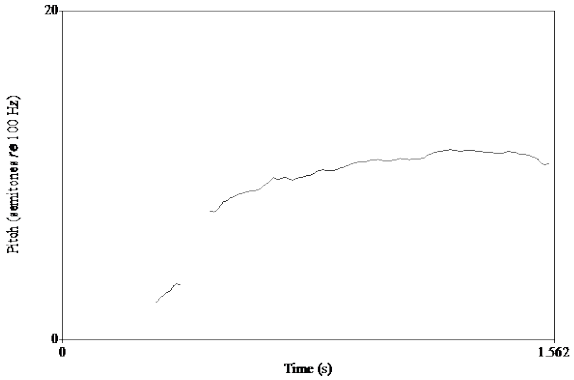


Fig.9: “dr”

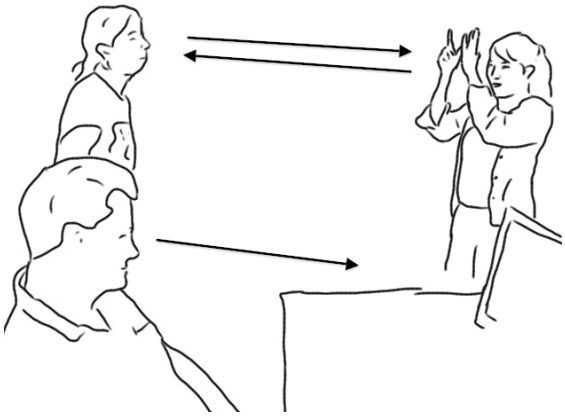
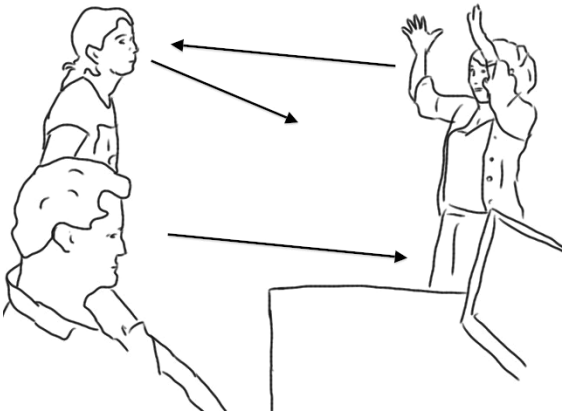


Fig.10: “-rrr-”



There is a transition relevance place (TRP) at the end of line (2), where the syntax and the action are complete, and intonation falls. The addressee, the head of the workshop who is responsible for building the props, has turned his gaze toward the computer and responds with a minimal *a* ’right’. However, the artist has lifted her hands (as shown in Figure 7) and continues with an embodied demonstration accompanied by a vocalization *drrrrrr*. The vocalization is done as a separate prosodic contour that iconically goes up at the beginning and falls slightly at the end (Figure 8). It follows the speaker’s hands that move upward, opening up, and then turning downward (Figures 9–10). The hands represent the fabric that moves from hanging position to plate-like structures when twirled, and back to hanging, in a kind of a “modeling gesture” (Enfield, 2009: 113–148). After the gesture + vocalization element is brought to a close, the speaker adds another clause, a new TCU. The embodied display thus constitutes an element quite similar to an incrementally added manner adverb at the end of the clause in (1–2), and the upcoming *ja* ‘and’-prefaced syntactic structure also builds on the display. However, it would reduce the specific accomplishment of the embodied element to force it into a regular syntactic analysis of what other materials could have occurred instead (after all, a separate clause is an option too in this turn-position). An embodied display is simply a qualitatively different kind of element that also employs turn-space in real time. In contrast to language, it is predominantly iconic, a depiction (Clark 2016). The body and the grammar are communicative tools used in parallel and sometimes interchangeably, the bodily movement partially gets its meaning from the verbal devices, and the other way around. Crucially for grammar, it is not adequate to analyze *ja sis tõmbub tgagasi nii* ‘and then retracts like this’ as if being coordinated with the previous clause *ta k- nagu tekib* ‘it appears’, because the embodied display is temporally,

prosodically, as well as content-wise a part of the emerging structure. Instead, we can propose multimodal patterns where embodied demonstrations are interwoven with grammar. Therefore analysts need to increase the range of tasks that grammar is said to handle, to include its capacity to project embodied elements and to incorporate them into the temporally emergent structure of speaker turns.

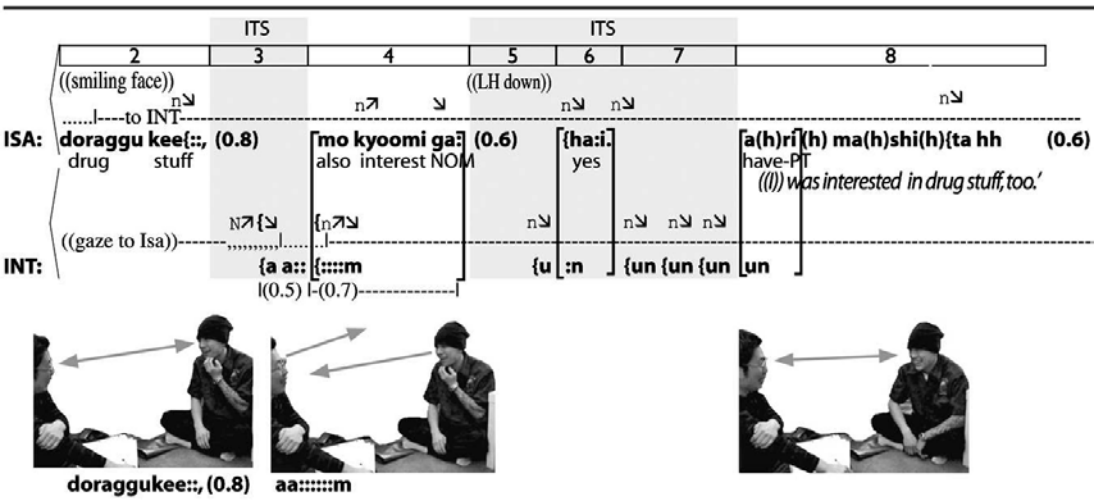
In summary, within the central area of turn construction we have relatively little knowledge about how exactly specific grammatical structures, such as conjunction/particle-initial TCUs, are contingent on bodily conduct by self and other. At the same time, we know that there is specific grammar for projecting embodied displays, depictions, and we are currently exploring the turn-constructional status of those that are syntactically relatively independent, such as shown in example (7). Depiction is a basic method of communication and regularly deployed in everyday interaction, while existing linguistic theories and processing models only account for descriptive methods (Clark 2016). The focus in this section has been on a single speaker turn and relatively straightforward syntax. We will now move into the area of bodily conduct features in the co-construction of exceedingly complex grammar.

4. Bodily conduct and emergent grammar

Close analysis of grammatical structure as emerging in interaction has revealed how it is flexibly adapted to local contextual contingencies. Legendarily, Charles Goodwin (1981) showed how a regular utterance, *I gave up smoking cigarettes one week ago today actually*, emerges across changing participation frameworks. He demonstrated how every next segment of the utterance is tailored to a different recipient, dependent on whether the recipient is knowing or unknowing about the reported event. The recipient's gaze, or the lack of it, essentially contributes to the emergence of the linguistic structure, which in this case also happens to constitute a full sentence. It was nevertheless neither planned, nor produced, as a single coherent proposition. Instead, the grammar of e.g. adding the adverbial phrase 'one week ago today' can be seen as a method of re-designing a prior utterance that has turned out to be inadequate for the new recipient. This basic insight by Goodwin, that recipients actively participate in producing grammatical structure, was recently complemented by Stoenica and Pekarek Doehler (2015), who demonstrated how relative clauses in French are systematically occasioned by the recipient's non-verbal display of trouble with referent identification. We can thus see how the bodily behavior of a recipient can have a crucial impact not only on the emerging syntactic unit in a speaker's turn, but also on the type of grammatical structure that emerges, such as a relative clause. Following in the footsteps of these path-breaking studies, we can begin to ask whether, or to what extent, responsiveness to local interactional concerns can constitute an alternative explanation of how grammar, i.e. patterned verbal behavior, is structured. It makes an essential difference in whether we see grammar as a product of the brain or as a method of dealing with interactional matters, such as turn-construction, recipient-design and affiliation. An obvious locus for searching for evidence of embodied behavior occasioning a certain kind of grammatical construction as a relevant next would be incrementally added structures, both clausal and phrasal ones, especially after pauses.

Studies on how "core" syntactic structures, such as noun phrases or predicates, emerge as a local reaction to embodied practices are still rare. A notable exception is Iwasaki's (2009, 2011) demonstration of how (clausal) units in Japanese emerge as a result of a dynamic interactive process, which includes "interactive turn spaces" – places where the speaker of the unit-in-progress invites the interlocutors to co-participate in the building of the action, with both the invitation and the responses being partially done through embodied conduct. Example (8) shows an instance where Isa tells about his reasons for moving to the US. He creates interactive turn spaces for his interlocutor to react after the two phrases, *doraggu kee* 'drug stuff' and *kyoomi* 'interest', before completing the clause with the verb *arimashita* 'had'. Gaze, head nods and gesture (transcribed above each speaker line) are involved in the two micro-collaborations during the production of a single clause. The pictures show the teller Isa to the right.

(8) Multimodal micro-collaboration within a clause: Iwasaki (2009: 240, 2011;
ITS = interactive turn space, 2–8 = numbers for analytic segments, n = nod)



Smiling, nodding and turning gaze to the recipient, Isa utters *doraggu kee* 'drug stuff' (in segment 2), which Int receives with a nod + response token during the interactive turn space (segment 3). The teller has thus explicitly secured recipiency for the continuation of the initiated syntactic structure. Somewhat more complex, the next segment *mo kyoomi ga* 'also interest' + nod by Isa (segment 4) is received by several response tokens and nods by Int (segments 5–7), and after the first of these, Isa also responds with an affirmation *hai* 'yes' (segment 6). He does so before actually completing his clause-under-production. In short, by using response particles and nodding Int actively participates in the emerging grammatical structure uttered by Isa. The structure itself is thus a result of several interactive episodes.

A few studies have targeted grammar in multimodal structures. It has, for example, been shown that a speaker can complete a grammatically incomplete structure with an explanatory gesture, which is useful in word searches (Hayashi 2003) and second language conversation (Olsher 2004, Mori & Hayashi 2006). These embodied displays effectively take the position of lexical units; not only do language and gesture complement each other when making meaning in so-called *composite utterances* (Enfield

2009) but an embodied demonstration occupies a grammatical and temporal slot within the emerging syntax (Keevallik 2013). A Swedish example from a dance class in Lindy Hop follows (9). One of the teachers is explaining a leading technique and the consequences of a wrong move for the followers ("ladies" in the transcript) in the dancing couple. The complex conditional structure in lines 4–5 involves two embodied demonstrations, each of which completes an initiated clause, *om killarna gör* 'if guys do', and *så kommer ju tjejerna att göra* 'then the ladies will do'. They constitute two separate syntactic-bodily units, where the demonstrations fill an essentially "syntactic" slot. Unfortunately, the elaborate dance demonstrations cannot be represented in a comprehensible manner through images.

(9) Syntactic-bodily units, syntactic initiation + demonstration; dance instruction (author's corpus)

```
1 Lead:    Och det är VÄLdigt väldigt väldigt viktigt. (0.2)
           and it  is very    very    very    important
           'And it is really really really important.'

2          hos dem   som FÖR. (.) initialt. (.) för     det  är
           for those who lead     initially     because it is
           'For those who lead. (.) Initially. (.) Because this'

3          DET  som   kommer att sprida sig på: TJEjer  också.
           this which FUT        spread REF to  lady-PL too
           'is what is going to be transferred to the ladies too.'

4          För    *om KILlarna   gör - (2.0)*
                  *d e m o n s t r a t i o n*
           because if guy-PL-DEF do
           'Because if the guys do'

5          Så  *kommer ju  TJEjerna     att göra - (1.3)*
               *  d e m o n s t r a t i o n            *
           then FUT    PAR lady-PL-DEF  to  do
           'then the ladies will do'

6          eftersom de   följer. Eller hur.
           because  they follow  or    how
           'because they follow. Right?'
```

After the second syntactic-bodily unit has come to a completion at the end of line 5, the turn can be continued with another subordinate clause, the causal 'because they follow'. The resulting structure involves a conditional clause-initiation + embodied demonstration (line 4), followed by another clause initiation + embodied demonstration (line 5), to which a causal clause is incrementally added (line 6), well-timed with the embodied demonstration, not the previous verbal segment. In this way embodied demonstrations participate in the temporal evolution of a complex multimodal structure, where the verbal segments on their own would be incomplete and incomprehensible. *Especially because the resulting structure makes perfect sense in the activity context, we should be systematically describing the grammatical opportunities for this kind of incorporation of the moving body into syntax*.

Syntax can be discontinued for an embodied demonstration after verbs, copulas and quotatives, but also adverbial phrases, adjectives, articles and subjects (Keevallik 2015).

In perhaps the most extreme cases, contrastive conjunctions and prepositions can be used in between two embodied demonstrations, indexing that the body is currently launching a contrasting action. For example, the Swedish compound preposition *istället för* 'instead of' as well as the English *instead of* are used to launch the incorrect performance in contrast to what has been going on so far, a demonstration of a correct posture (Keevallik, in press). Thus, a preposition does not necessarily project a noun phrase, as a verbal-only grammatical account might suggest. The study concluded that functionally the above prepositions instead belong together with contrastive conjunctions, such as *but*. Paradigms extracted from verbal-only activities do not necessarily fit the systematicity we find in multimodal interaction.

As was discussed in example (9) embodied demonstrations complete linguistic projections. It is likewise possible to retrospectively treat an embodied demonstration as an element in the emerging unit, basically a clause (Keevallik 2013). This yet again calls into question the analytical boundary between verbal and bodily conduct, and illustrates that there is no reason to limit the syntactic-bodily units to the "verbal first – embodied later" pattern. A case in point is shown in example (10). Here the Swedish dance class is working on a longer step sequence and has just practiced the first section accompanied by the teacher's vocalization *di:gida gidi:gida* – a non-lexical accompaniment common in this kind of dance practice (Keevallik 2015). It has been made clear that there are some steps left after the students have mastered this part. In the excerpt, the teacher once again guides the students through the *di:gida gidi:gida* (in line 2) but then adds the next segment of the step sequence accompanied by *wak:ido*. After *wak:ido* the teacher completes the clause, incorporating the *wak:ido* as if it was an initially positioned object in line (3). Importantly, the ad-hoc sounds *wa, k:i,* and *do* do not constitute a word in any sense, and are instead indexically tied to the teacher's simultaneous body movements. Nevertheless without the [embodied demonstration + *wak:ido*] the clause *lägger vi till nu*, lit. 'put we on/to now' would lack a grammatically obligatory element. In short, the multimodal structure emerges in this activity context as a combination of an embodied demonstration, followed by an instruction. The multimodal "clause" in this combination of actions can again only be identified retrospectively.

(10) Demonstration + syntactic completion; dance instruction (Keevallik 2013)

```
1 Follow:  En  gång till. Fem  sex sju   å.
           one time more  five six seven and
           'Once more. Five six seven and'

2          *wa, (.) pa, (.) di:gida gidigiDA, (.)*
           *   j o i n t     p r a c t i c e     *

3          *wak:iDO,* (0.2) lägger vi till  nu. (0.2)
           *  demo  *
                          put    we on/to now
                          'will be added now.'

4          Då   blir   det *wah pa (.) di:gida gidi:gida, (.) wap:iDO.*
                           *  d  e  m  o  n  s  t  r  a  t  i  o  n   *
           then be:FUT it
           'Then it'll be'
```

Finally, the ultimate evidence of syntactic-bodily units being real for the participants is the occurrence of collaborative completion; a verbal initiation by one participant can be fulfilled by another who performs an embodied demonstration (Keevallik 2013, see also Hayashi (2005) for joint turn construction through language and the body). Grammar thus emerges – or ceases to emerge – in a way that is sensitive to other participants' verbal or embodied actions, fitted to local contingencies. This section has shown that this is also true when it comes to core grammar, such as clauses, and complex syntactic structures involving coordinate and subordinate clauses. As a different kind of example, a speaker may abandon a TCU-in-progress when the action has been treated as complete through the recipient's bodily-visual response (Ford et al. 2012: 206). Most crucially, examples (9–10) evidence the participants' ability to incrementally build grammatical constructions by interweaving bodily conduct seamlessly into linguistic structure. A grammar that includes a systematic description of these patterns and their information structure would reveal more realistically how language is one among the many sense-making devices in human interaction. *The emergence of syntactic-bodily units illustrates why grammar cannot be strictly kept apart from embodied behavior*.

5. Conclusion

Research on interaction helps us realize that grammar and lexis only make sense in local multimodal ecologies; grammar emerges incrementally and responsively to interactional contingencies, among them embodied ones. Information and meaning are not only channeled into verbal mode. Grammar cannot be extracted from the flow of embodied interaction without undermining our understanding of what it actually is and does. Gestures, eye gaze, head movement, body position and prosody are visually and/or auditorily available to recipients, and thus partly constitute the meaning of the utterance (Fox 2007: 312). Likewise, the temporal emergence of structure in turns-at-talk, where verbal elements are deployed interchangeably with, and in response to, embodied displays should be a central concern when discussing grammar. We need to account for how people combine various modalities when speaking to each other.

The demonstrating and depictive capacities of language should be acknowledged alongside descriptive ones. Partially, the different modalities may serve different purposes (e.g. Stevanovic and Monzoni 2016). Verbal resources are mostly symbolic whereas bodily ones often iconic and indexical, and thus complementary. As the current overview has shown, transitions between the two are not uncommon, and the modalities not hermetically separated when it comes to building sequences, actions and units, including units described in regular grammars. Among other things, it was argued that embodiments should at least sometimes be included in the analysis of syntactic structure, such as a clause, prepositional phrase, or coordination, even though we may need different terminology to handle e.g. a [preposition + embodied demonstration], if we do not want to call it a "phrase". In any case, realistic abstractions of grammatical constructions should not be that distant from actual utterances.

In the above we moved through multimodal sequences to multimodal turns. Embodied action can be positioned and treated as part of a conversational sequence, both as a first or second action, after which the next relevant action is grammatically formatted in a manner that is appropriate to that sequential position. It was also discussed how

grammar may emerge in response to local interactional matters, such as a puzzled gaze occasioning a relative clause, which leads us to wonder how these matters could be incorporated into the systematic understanding of grammar. Grammar accomplishes sequential actions. We could therefore pose the question the other way around and ask how embodied local practices may have sedimented over time into linguistic patterns, and how various non-verbal or depictive segments might systematically figure in these patterns, for example in the form of syntactic-bodily units. Occasionally, grammar seems to emerge in different paradigms and syntagms than what has been found in activities where verbal modality dominates. For example, devices used for contrasting embodied demonstrations are organized somewhat differently from those used for contrasting clauses. To focus only on the verbal aspects of language use is to miss its flexibility and applicability to a broad range of communicative tasks.

Linguistic traditions that have deliberately excluded complementary modes of communication, or even the entire context, from the analysis, would be clearly found lacking when accounting for sense-making in the activity environments described above. As Mondada (2014b: 139) puts it: "Multimodal resources are integrated in a holistic way and make sense together; moreover, they can be seen as not being a priori hierarchized, but as having their relevance empirically and situatedly defined within the context of the activity and its ecology." It depends on the activity whether the verbal mode is a central meaning-making device or not. In primarily embodied actions, such as offering and requesting a concrete object, language may in fact have a subordinate role and feature numerous fixed formats, such as a 'here you are', 'please have a look', etc. In these actions we are able to see an important aspect of grammar – that of being merely one building-block of a complex bodily-verbal (cultural) practice. In contrast, primarily conversational activities are more likely to feature complex grammar and detachment, i.e. talk about matters beyond the here-and-now. These contexts have been the focus of analysis since Saussure, because they provide ample examples of linguistic structure. However, if we are interested in documenting the full range of grammatical function, we need to broaden our understanding of what constitutes the appropriate data for analysis.

An especially versatile locus for studying coordination between grammar and the body seems to be instructional activities, where the objective is to teach another's body to perform a task. In these settings, we are likely to find a prolific use of both embodied and grammatical devices of sense-making. It is here we can frequently locate syntactic-bodily units in which an embodied demonstration literally takes the temporal slot of an element within the emerging grammatical structure. These patterns could be included in a systematic description of grammar. They will also change our conceptualization of what grammar is: a method of accomplishing social action in fine coordination with embodied behavior, a structure that emerges across participants in multimodal interaction.

Transcription signs

CAPITAL LETTERS	– emphasis
[ ]	– overlap
*	– simultaneity

=	– latching of turns or words
(0.5)	– pause length in tenths of a second
(.)	– micropause
:	– lengthening of a sound
° soft °	– soft voice
.h	– breathing in
(there)	– the item in the translation does not occur in the original
((snort))	– transcriber's comments and descriptions
.	– falling intonation at the end of the intonation unit
?	– rising intonation at the end of the intonation unit
,	– non-final intonation at the end of the intonation unit
[space] –	– level intonation
↑	– markedly high rise in pitch
#	– number of the picture

DEF	– definite form
FUT	– future
INF	– infinitive
NOM	– nominative
NPST	– non-past
PAR	– particle
PL	– plural
REF	– reflexive
SG	– singular
1,2,3	– person

References

Arminen, I., Koskela, I., & Palukka, H. (2014). Multimodal production of second pair parts in air traffic control training. *Journal of Pragmatics*, 65, 46–62. doi:10.1016/j.pragma.2014.01.004

Broth, M. (2014). Pans, Tilts and Zooms. Conventional Camera Gestures in TV-production. In M. Broth, E, Laurier & L. Mondada (Eds.), *Studies of Video Practices: Video at Work* (pp. 63–96). New York: Routledge.

Broth, M., & Keevallik, L. (2014). Getting ready to move as a couple. Accomplishing mobile formations in a dance class. *Space and Culture,* 17(2), 107–121. 10.1177/1206331213508483

Bybee, J. 2003. Cognitive processes in grammaticalization. In M. Tomasello (Ed.), *The New Psychology of Language, Volume II* (pp. 145–167). Mahwah: Erlbaum.

Cekaite, A. (2010). Shepherding the child: embodied directive sequences in parent-child interactions. *Text and Talk,* 30(1), 1–25. https://doi.org/10.1515/text.2010.001

Clark, H.H. (2016). Depicting as a method of communication. *Psychological Review,* 123(3), 324–347. http://dx.doi.org/10.1037/rev0000026

Clift, R. (2016). *Conversation Analysis*. Cambridge: Cambridge University Press.

Couper-Kuhlen, E., & Selting, M. (1996). *Prosody in Conversation: Interactional Studies*. Cambridge: Cambridge University Press.

Couper-Kuhlen, E., & Selting, M. (To appear). *Interactional Linguistics: Studying Language in Social Interaction*. Cambridge: Cambridge University Press.

Deppermann, A. (2013a). Introduction: Multimodal interaction from a conversation analytic perspective. *Journal of Pragmatics*, 46(1), 1–7. doi:10.1016/j.pragma.2012.11.014

Deppermann, A. (2013b). Turn-design at turn-beginnings: Multimodal resources to deal with tasks of turn-construction in German. *Journal of Pragmatics*, 46(1), 91–121. http://dx.doi.org/10.1016/j.pragma.2012.07.010

De Stefani, E., & Gazin, A.-D. (2014). Instructional sequences in driving lessons: Mobile participants and the temporal and sequential organization of actions. *Journal of Pragmatics*, 65, 63–79. doi:10.1016/j.pragma.2013.08.020

Enfield, N. (2009). *The anatomy of meaning: Speech, gesture, and composite utterances*. Cambridge, UK: Cambridge University Press.

Eriksson, M. (2009). Referring as interaction: On the interplay between linguistic and bodily practices, *Journal of Pragmatics*, 41(2), 240–262. http://dx.doi.org/10.1016/j.pragma.2008.10.011

Ford, C.E., Thompson, S.A., & Drake, V. (2012). Bodily-visual practices and turn-continuation. *Discourse Processes,* 49(3–4), 192–212. 10.1080/0163853X.2012.654761

Fox, B. (2007). Principles shaping grammatical practices: An exploration. *Discourse Studies,* 9(3), 299–318. https://doi.org/10.1177/1461445607076201

Fox, B., & Heinemann, T. (2015). The alignment of manual and verbal displays in requests for the repair of an object. *Research on Language and Social Interaction,* 48(3), 342–362. 10.1080/08351813.2015.1058608

Fox, B.A., & Robles, J. (2010). *It's like mmm*: Enactments with *it's like*. *Discourse Studies*, 12(6), 715–738. https://doi.org/10.1177/1461445610381862

Gallagher, S. (2011). Interpretations of embodied cognition. In W. Tschacher & C. Bergomi (Eds.), *The Implications of Embodiment: Cognition and Communication* (pp. 59–71). Exeter: Imprint Academic.

Ginzburg, J., & Poesio, M. (2016). Grammar is a system that characterizes talk in interaction. *Frontiers in Psychology*, 7, 1938. DOI=10.3389/fpsyg.2016.01938

Goffman, E. (1981). *Forms of Talk*. Oxford: Basil Blackwell.

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'. *Journal of Pragmatics*, 32(1), 24–59. http://dx.doi.org/10.1016/S0378-2166(99)00030-2

Goodwin, C. (1980). Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning. *Sociological Inquiry*, 50(3–4), 272–302. 10.1111/soin.1980.50.issue-3-4

Goodwin, C. (1981). *Conversational Organization: Interaction between Speakers and Hearers*. New York, NY: Academic Press.

Goodwin, M.H., & Goodwin, C. (1987). Children's Arguing. In S. Philips, S. Steele, & C. Tanz (Eds.), *Language, Gender, and Sex in Comparative Perspective* (pp. 200–48). Cambridge: Cambridge University Press.

Goodwin, C., & Goodwin, M.H. (1992). Assessments and the Construction of Context. In A. Duranti, & C. Goodwin (Eds.), *Rethinking Context: Language as an Interactive Phenomenon* (pp. 147–90). Cambridge: Cambridge University Press.

Goodwin, C., (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32, 1489–1522. http://dx.doi.org/10.1016/S0378-2166(99)00096-X

Hayashi, M. (2003). Language and the body as resources for collaborative action: A study of word searches in Japanese conversation. *Research on Language and Social Interaction*, 36(2), 109–141. http://dx.doi.org/10.1207/S15327973RLSI3602_2

Hayashi, M. (2005). Joint turn construction through language and the body: Notes on Embodiment in coordinated participation in situated activities. *Semiotica,* 156, 21–53. https://doi.org/10.1515/semi.2005.2005.156.21

Hopper, P. (1998). Emergent grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 155–175). Mahwah, NJ: Lawrence Erlbaum.

Iwasaki, S. (2009). Initiating interactive turn spaces in Japanese conversation: local projection and collaborative action, *Discourse Processes*, 46(2), 226–246. http://dx.doi.org/10.1080/01638530902728918

Iwasaki, S. (2011). The multimodal mechanics of collaborative unit construction in Japanese conversation. In J. Streeck, C. Goodwin & C. LeBaron (Eds.), *Embodied Interaction: Language and Body in the Material World* (pp. 106–120). Cambridge, UK: Cambridge University Press.

Keevallik, L. (2012). Compromising progressivity: 'no'-prefacing in Estonian. *Pragmatics,* 22(1), 119–146. 10.1075/prag.22.1.05kee

Keevallik, L. (2013). The interdependence of bodily demonstrations and clausal syntax. *Research on Language and Social Interaction,* 46(1), 1–21. http://dx.doi.org/10.1080/08351813.2013.753710

Keevallik, L. (2014). Turn organization and bodily-vocal demonstrations. *Journal of Pragmatics,* 65, 103–120. http://dx.doi.org/10.1016/j.pragma.2014.01.008

Keevallik, L. (2015). Coordinating the temporalities of talk and dance. In A. Deppermann & S. Günthner (Eds.), *Temporality in Interaction* (pp. 309–336). Amsterdam, Philadelphia: John Benjamins.

Keevallik, L. (2016). Affirmative particles as response cries. Plenary talk at OFTI 34, Helsinki 15–16.09.2016

Keevallik, L. (in press): Linking performances: The temporality of contrastive grammar. In E. Couper-Kuhlen & R. Laury (Eds.), *Linking in Grammar and Action* (pp. xx-xx). Helsinki: Finnish Literary Society.

Keisanen, T. (2012). "Uh-oh we were going there": Environmentally occasioned noticings of trouble in in-car interaction. *Semiotica,* 191(1/4), 199–124. https://doi.org/10.1515/sem-2012-0061

Keisanen, T., & Kärkkäinen, E. (2014). A multimodal analysis of compliment sequences in everyday English interactions. *Pragmatics*, 24(3), 649–672. 10.1075/prag.24.3.09kei

Keisanen, T., & Rauniomaa, M. (2012). The organization of participation and contingency in pre-beginnings of request sequences. *Research on Language and Social Interaction,* 45(4), 323–351. http://dx.doi.org/10.1080/08351813.2012.724985

Kendrick, K.H., & Drew, P. (2016). Recruitment: Offers, requests, and the organization of assistance in interaction. *Research on Language and Social Interaction*, 49(1), 1–19. http://dx.doi.org/10.1080/08351813.2016.1126436

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge, UK: Cambridge University Press.

Kärkkäinen, E., & Keisanen, T. (2012). Linguistic and embodied formats for making (concrete) offers. *Discourse Studies*, 14(5), 1–25. https://doi.org/10.1177/1461445612454069

Lindström, A., & Mondada, L. (Eds.) (2009). Special issue on Assessments in Social Interaction. *Research on Language and Social Interaction,* 42(4). http://dx.doi.org/10.1080/08351810903296457

Linell, P. (2009). *Rethinking Language, Mind, and World Dialogically: Interactional and Contextual Theories of Human Sense-Making*. Charlotte, NC: Information Age Publishing.

McNeill, D. (2005). *Gesture and thought*. Chicago: Chicago University Press.

Mondada, L. (2007). Multimodal resources for turn-taking: pointing and the emergence of possible next speakers. *Discourse Studies,* 9, 194–225. https://doi.org/10.1177/1461445607075346

Mondada, L. (2009a). Emergent focused interactions in public places: A systematic analysis of the multimodal achievement of a common interactional space. *Journal of Pragmatics*, 41 (10), 1977–1997. http://dx.doi.org/10.1016/j.pragma.2008.09.019

Mondada, L. (2009b). The embodied and negotiated production of assessments in instructed actions. *Research on Language and Social Interaction,* 42(4), 329–361. http://dx.doi.org/10.1080/08351810903296473

Mondada, L. (2014a). Requesting immediate action in the surgical operating room: Time, embodied resources and praxeological embeddedness. In P.Drew & E. Couper-Kuhlen (Eds.), *Requesting in Social Interaction* (pp. 269–302). Amsterdam, Philadelphia: John Benjamins.

Mondada, L. (2014b). The local constitution of multimodal resources for social interaction. *Journal of Pragmatics,* 65, 137–156. http://edoc.unibas.ch/dok/A6328898

Mori, J., & Hayashi, M. (2006). The achievement of intersubjectivity through embodied completions: A study of interactions between first and second language speakers. *Applied Linguistics,* 27, 195–219. https://doi.org/10.1093/applin/aml014

Mortensen, K. (2009). Establishing recipiency in pre-beginning position in the second language classroom. *Discourse Processes,* 46, 491–515. 10.1080/01638530902959463

Mortensen, K. (2016). The body as a resource for other-initiation of repair: Cupping the hand behind the ear. *Research on Language and Social Interaction,* 49(1), 34–57. http://dx.doi.org/10.1080/08351813.2016.1126450

Musk, N. (2016). Correcting spellings in second language learner's computer-assisted collaborative writing. *Classroom Discourse*, 7, 36–57. http://dx.doi.org/10.1080/19463014.2015.1095106

Nevile, M. (2015). The embodied turn in research on language and social interaction. *Research on Language and Social Interaction,* 48(2), 121–151. 10.1080/08351813.2015.1025499

Olsher, D. (2004). Talk and gesture: The embodied completion of sequential actions in spoken interaction. In R. Gardner & J. Wagner (Eds.), *Second language conversations* (pp. 221–245). London, England: Continuum.

Ochs, E., Schegloff, E.A., & Thompson, S.A. (Eds.) (1996). *Interaction and Grammar*. Cambridge: Cambridge University Press.

Pinker, S. (1995). *The Language Instinct: How the Mind Creates Language*. New York, NY: HarperPerennial

Rasmussen, G., Hazel, S., & Mortensen, K. (Eds.) (2014). *A body of resources – CA studies of social conduct*. Special Issue for *Journal of Pragmatics*, *65*.

Rauniomaa, M., & Keisanen, T. (2012). Two multimodal formats for responding to requests. *Journal of Pragmatics,* 44(6–7), 829–842. http://dx.doi.org/10.1016/j.pragma.2012.03.003

Rossi, G. (2012). Bilateral and unilateral requests: The use of imperatives and *Mi X*? interrogatives in Italian. *Discourse Processes*, 49(5), 426–458. http://dx.doi.org/10.1080/0163853X.2012.684136

Rossi, G. (2014). When do people not use language to make requests? In P.Drew & E. Couper-Kuhlen (Eds.), *Requesting in Social Interaction* (pp. 303–334). Amsterdam, Philadelphia: John Benjamins.

Ruusuvuori, J. & Peräkylä, A. (2009). Facial and verbal expressions in assessing stories and topics. *Research on Language and Social Interaction,* 42(4), 377–394. http://dx.doi.org/10.1080/08351810903296499

Sacks, H., Schegloff, E.A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language,* 50, 696–735. 10.2307/412243

Schegloff, E.A. (1984). On Some Gestures' Relation to Talk. In J. M. Atkinson & J. Heritage, (Eds.), *Structures of Social Action: Studies in Conversation Analysis* (pp. 266–295), Cambridge: Cambridge University Press.

Schegloff, E.A. (1996). Turn organization: one intersection of grammar and interaction. In E. Ochs, E.A. Schegloff & S.A. Thompson (Eds.), *Interaction and grammar* (pp. 52–133). Cambridge: Cambridge University Press.

Schegloff, E.A. (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge: Cambridge University Press.

Selting, M., & Couper-Kuhlen, E. (2001). *Studies in Interactional Linguistics*. Amsterdam, Philadelphia: John Benjamins.

Seo, M-S., & Koshik, I. (2010). A conversation analytic study of gestures that engender repair in ESL conversational tutoring. *Journal of Pragmatics,* 42(8), 2219–2239. http://dx.doi.org/10.1016/j.pragma.2010.01.021

Sidnell, J. (2010). *Conversation Analysis: An Introduction*. Malden: Wiley-Blackwell.

Sinclair, J., & Coulthard, M. (1975) *Towards an Analysis of Discourse*. Oxford: Oxford University Press.

Sorjonen, M.-L., & Raevaara, L. (2014). On the grammatical form of requests at the convenience store. In P. Drew & E. Couper-Kuhlen (Eds.), *Requesting in Social Interaction* (pp. 243–268). Amsterdam, Philadelphia: John Benjamins.

Stevanovic, M. & Monzoni, C. (2016). On the hierarchy of interactional resources: Embodied and verbal behavior in the management of joint activities with material objects. *Journal of Pragmatics,* 103, 15–32. http://dx.doi.org/10.1016/j.pragma.2016.07.004

Stivers, T. (2008). Stance, alignment and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction,* 41(1), 31–57. http://dx.doi.org/10.1080/08351810701691123

Stivers, T. & Sidnell, J. (2005). Multi-modal interaction. *Semiotica,* 156, 1–20. 10.1515/semi.2005.2005.156.1

Stoenica, I.-M., & Pekarek Doehler, S. (2015). Relative clauses as turn continuations in French talk-in-interaction. Paper presented at the 14th International Pragmatics Conference, Antwerp, 26–31 July 2015.

Streeck, J. (2002). Grammars, words, and embodied meanings: on the uses and evolution of *so* and *like*. *Journal of Communication,* 52, 581–596. 10.1111/j.1460-2466.2002.tb02563.x

Streeck, J. (2009). *Gesturecraft: The Manu-Facture of Meaning*. Amsterdam: John Benjamins.

Streeck, J., Goodwin, C., & LeBaron, C. (Eds.) (2011). *Embodied Interaction: Language and Body in the Material World.* Cambridge: Cambridge University Press

Tulbert, E., & Goodwin, M. H. (2011). Choreographies of Attention: Multimodality in a. Routine Family Activity. In J. Streeck, C. Goodwin, & C. LeBaron (Eds.), *Embodied Interaction: Language and Body in the Material World* (pp. 79–92). Cambridge: Cambridge University Press.

Walsh, S. (2011). *Exploring Classroom Discourse: Language in Action*. London, New York: Routledge.