RESEARCH ON LANGUAGE AND SOCIAL INTERACTION
2021, VOL. 54, NO. 2, 183–202
https://doi.org/10.1080/08351813.2021.1899710

The Routinization of Grammar as a Social Action Format: A Longitudinal Study of Video-Mediated Interactions

Simona Pekarek Doehler [a] and Ufuk Balaman [b]

[a]Center for Applied Linguistics, Institute of Language Sciences, University of Neuchâtel, Switzerland; [b]Department of English Language Teaching & HUMAN Research Centre, Hacettepe University, Turkey

ABSTRACT

In this article, we provide longitudinal evidence for the progressive routinization of a grammatical construction used for social coordination purposes in a highly specialized activity context: task-oriented video-mediated interactions. We focus on the methodic ways in which, over the course of 4 years, a second language speaker and initially novice to such interactions coordinates the transition between interacting with her coparticipants and consulting her own screen, which suspends talk, without creating trouble due to halts in progressivity. Initially drawing on diverse resources, she increasingly resorts to the use of a prospective alert constructed around the verb *to check* (e.g., "I will check"), which eventually routinizes in the lexically specific form "let me check" as a highly context- and activity-bound social action format. We discuss how such change over the participant's video-mediated interactional history contributes to our understanding of social coordination in video-mediated interaction and of participants' recalibrating their grammar-for-interaction while adapting to new situations, languages, or media. Data are in English.

People coordinate their social interaction with others through the use of mutually recognizable practical procedures for action—that is, "methods" in the ethnomethodological sense of the term (Garfinkel, 1967). Based on these procedures, they manage fundamental organizational issues such as turn-taking, sequencing of actions, and opening or closing conversation. Methods develop experientially over time, based on people's repeated engagement in locally accomplished social interactions (Pekarek Doehler et al., 2018), and so do the multisemiotic resources deployed as part of these methods. Methods and the related resources, the sharedness of which warrants mutual accountability of social actions (Garfinkel, 1967), constitute people's interactional competence.

Existing longitudinal studies on the development of interactional competence in a first (L1) or a second/foreign/additional language (L2) across a range of different settings converge on evidencing that such development implies the progressive diversification of practices and resources for interaction (on L1 see, e.g., Wootton, 1997; on L2 see overviews by Pekarek Doehler & Pochon-Berger, 2015; Skogmyr Marian & Balaman, 2018). This diversification allows for increasingly context-sensitive and recipient-designed conduct (cf. Sacks & Schegloff, 1979; Sacks et al., 1974)—that is, conduct that is more and more acutely tailored to the precise situation and the precise others at hand (see also Skogmyr Marian, 2021/this issue).

Such development is not just part of L1 or L2 learning but is constitutive of the very way we navigate our social lives. Our interactional histories bring about unprecedented types of social situations, which we enter as novices, progressively adapting our practices to the precise constraints

CONTACT Simona Pekarek Doehler simona.pekarek@unine.ch Center for Applied Linguistics, FLSH, University of Neuchâtel, Pierre-à-Mazel 7, Neuchâtel CH-2001, Switzerland.

We thank Arnulf Deppermann and two anonymous reviewers for their detailed comments on a previous version of this paper.

and affordances of the situations at hand (Deppermann, 2018) while we move toward becoming more "competent members" (Goodwin, 2018). One such situation is task-based video-mediated interaction (VMI). Navigating the complex digital-social ecology (Luff et al., 2003) of VMI requires, on the part of the participants, highly context- and medium-specific coordination work that is consequential for the maintenance of progressivity and intersubjectivity. Yet little is known about how novices chart their way through the related medium-specific interactional exigencies (but see Balaman, 2018; Balaman & Sert, 2017a, 2017b). Also, although we have solid evidence for how linguistic resources are put to use for the specific purpose of coordinating social interaction (Couper-Kuhlen & Selting, 2018), we know only little about how people recalibrate their grammar-for-interaction over time as part of their adapting to new situations, languages, or media.

The present study sets out to address these two gaps: It investigates how, over the course of four years, a participant—L2 speaker and initially novice to task-oriented VMI—goes about navigating the interface between her social interaction with the remote coparticipants and her own screen-based activity, which typically disrupts the progressivity of talk-in-interaction (Balaman, 2018, 2019; Balaman & Sert, 2017a). We show how she shifts between multiple activities (Haddington et al., 2014) in methodic ways, deploying a range of practices through which she preemptively accounts for the suspension of her own talk and bids for a temporary halt of others' talk, thereby successfully initiating suspension of the turn-taking machinery so as to fully focus on her screen-based searches. Initially, she uses a variety of resources for doing so, yet over time, she increasingly draws on a grammatical schema constructed around the verb *to check* ("I will check X"), which eventually routinizes in the lexically specific form "let me check" as a highly context- and activity-bound resource for such coordination. By the same token, her practices move from being implemented as informings about her future doing ("I will X") to formally shaping up as request for permission to do some action ("let me X").

The study sheds further light on the multilayeredness of VMI as a form of cooperative (Goodwin, 2018) action (Arminen, Licoppe & Spagnolli, 2016; Licoppe & Morel, 2014); it also offers results that stand in sharp contrast to established, typically cognitively oriented understandings of learning and development: Rather than identifying diversification of practices over time or an increase in grammatical complexity, we document the progressive streamlining of a single lexically specific construction as highly context- and activity-bound "social action format" (Fox, 2007). This provides evidence for the development of grammar-for-interaction as an integral part of the development of the speaker's ability to interact with others and highlights the fact that such development is not limited to having more forms but centrally involves the ability to deploy practical solutions for getting locally relevant interactional work done.

Background

Video-mediated interactions in task-oriented settings

In online collaborative task-oriented settings, the geographically dispersed participants, interacting through Google Hangouts, Skype, and the like, deploy screen-based activities for retrieving information online, yet they are also normatively expected to attend to the progressivity of talk-in-interaction (cf. Heritage, 2007) so as to avoid delaying task completion and disrupting intersubjectivity (Balaman, 2018; Balaman & Sert, 2017a). Although the screen-based searches are instrumental for the accomplishment of the joint task, they typically entail lengthy silences (Näslund, 2016). If these silences are not made recognizable as part of the joint task-oriented activity, they represent a potential source of interactional trouble, as they typically suspend the progressivity of talk-in-interaction (cf. Rintel, 2013).

In the data under scrutiny, the potential of disruption and trouble is augmented by the mutual nonaccessibility (Heath & Luff, 1993, 2000) of participants' individual screen activities. Due to the distortedness of gaze behaviors caused by the interactional asymmetry in VMI (Heath & Luff, 1993, 2000), participants are not able to identify when and for what purpose coparticipants attend to their screens. Differently from "ordinary" Skype conversations, in which participants orient toward the

"maxim 'put the face of the current speaker on screen', so that a 'talking heads' configuration is the default interaction mode" (Licoppe & Morel, 2014, p. 138, 2012), in our task-oriented VMIs participants maximize the visibility of screen documents while minimizing the visibility of their mutual heads. Therefore, when retrieving information through the screen, participants recurrently engage in a concurrent course of action (screen-based activity) that may suspend talk *in ways that are not accountable to coparticipants* (Hjulstad, 2016; Luff et al., 2016; Whalen & Zimmerman, 1998). They are hence faced with a practical problem of social coordination: how to orchestrate concurrent screen-based activity and talk-in-interaction without causing interactional trouble.

We show that this coordination work is methodically achieved based on a range of practices through which individual participants alert coparticipants to their incipient screen-based activities, thereby preemptively providing an account for their own cessation of talk while at the same time offering a bid for the suspension of the turn-taking machinery.

Social action formats: Grammar-for-interaction and its development over time

Grammar is a central part of participants' methods for coordinating social interaction (Couper-Kuhlen & Selting, 2018; Ochs et al., 1996) and building actions in mutually accountable ways. Yet how novices to new situations (including L2 speakers) recalibrate their grammars across their interactional histories as resources for interaction remains largely unexplored.

The development of grammar-for-interaction has only recently started to be addressed in longitudinal studies on L2 interactions. Eskildsen (2011) showed how an L2 English speaker in classroom interactions used the multiword expression *what do you say* first exclusively as a solicit for help within word-search sequences but later employed it also as a repetition request or a solicit for personal opinion. Pekarek Doehler (2018) documented the development of an "L2 grammar-for-interaction," showing how L2 speakers of French move from a "literal" use of the multiword expression *je sais pas* ("I don't know") as a claim of lack of knowledge to particle-like uses for opting out of turn and sequence. The developmental trajectories described by these studies show that the initially "literal" uses of multiword expressions become progressively complemented by interaction-organizational uses. Importantly, and distinct from other developmental studies, these results are not about the acquisition of "correct" grammar nor of stable form-function mappings but about the development of "social action formats," i.e., grammatical usage patterns specialized for accomplishing and coordinating actions in locally contingent ways (Fox, 2007).

Such recalibration of grammar-for-interaction is not restricted to language learning situations but is an integral part of how we move through our social lives, enter new situations, and adapt to new coparticipants. This has been demonstrated by Deppermann (2018), who documents how driving instructors' instructions to students become increasingly shorter, syntactically less complex, and elliptical as part of processes of mutual adaptation over time. Drawing on the aforementioned work, the present study follows a novice participant through her task-oriented VMI history. Although the novice participant in our study is a second-language speaker, we see the observed change in her practices as pertaining to the larger process of how people recalibrate their resources for the accomplishment of social actions when adapting to new situations, languages, and/or media.

Data and procedures

We explore longitudinal change in how one focal participant, SIN, coordinates her screen-based activities with talk in task-oriented VMI based on two data sets: DS1 (70 hours of interactions among groups of four participants meeting online once a week during 18 weeks) and DS2 (14 hours of dyadic interactions, collected four years later over a period of three weeks). All participants gave written consent for use and publication of the data for scientific purposes. We focus on SIN, an L2 English speaker whose L1 is Turkish, because she has participated in both data collection periods, which allows for tracking her practices over an extended amount of time, and because at the start of the recordings she is a novice in this highly specialized type of interaction. She has also reported not having

participated in any task-oriented VMI during the time between the two recording periods. Therefore, the data under scrutiny include all instances of SIN's engagement in task-oriented VMI during the targeted four-year period.

The online tasks in both data sets were part of an L2 English teacher education program. Such task-oriented VMIs represent a widespread educational device for enhancing students' L2 communicative practice. The tasks consisted of hinting and guessing episodes, e.g., guessing the name of famous inventions (Ex. 1), TV shows (Ex. 2 and 4), the content of a specific tweet (Ex. 3), or a movie (Ex. 5), all of which led the participants to undertake screen-based activities so as to retrieve various types of information for the benefit of collaborative task completion. The tasks hence entailed specific constraints on participation and collaboration, with ensuing consequences on such issues as turn-taking and the management of silence. Figure 1 shows a sample capture of the focal participant's screen. It illustrates how participants' limited visual access to each other was provided through a small video frame and exemplifies the multisemiotic resources participants selectively attended to while orienting to their screens, including text and images on Google websites or further documents. For both data sets, the screen-recording software operated in the background on each participant's computer separately; it captured talk, screen-based activities, and computer sounds.

In both data sets, we identified the need for participants to alert each other to their incipient screen-based activities as a practical issue they recurrently attend to. Participants orient to the "exclusive orders" (Mondada, 2014, p. 64) of their multiactivity, in which one activity (talk) is typically suspended in favor of another (screen-based search). We established, across both data sets, an initial inventory of all instances of SIN publicly displaying her incipient orientation to her own screen (including explicit verbal alerts, nonlexical vocalizations such as *hmmm*, and the use of grammatical projection to "buy time"). Bodily-visual conduct was found to be nonrelevant, likely due to the limited visual accessibility of participants (to each other as well as to us researchers) through the small skype frames (see Figure 1), which were often covered by documents or websites that participants were consulting on their respective screens.

To maximize comparability across occurrences and time (on methodological challenges for longitudinal CA research, see Depperman & Pekarek Doehler, 2021/this issue; Wagner et al.,

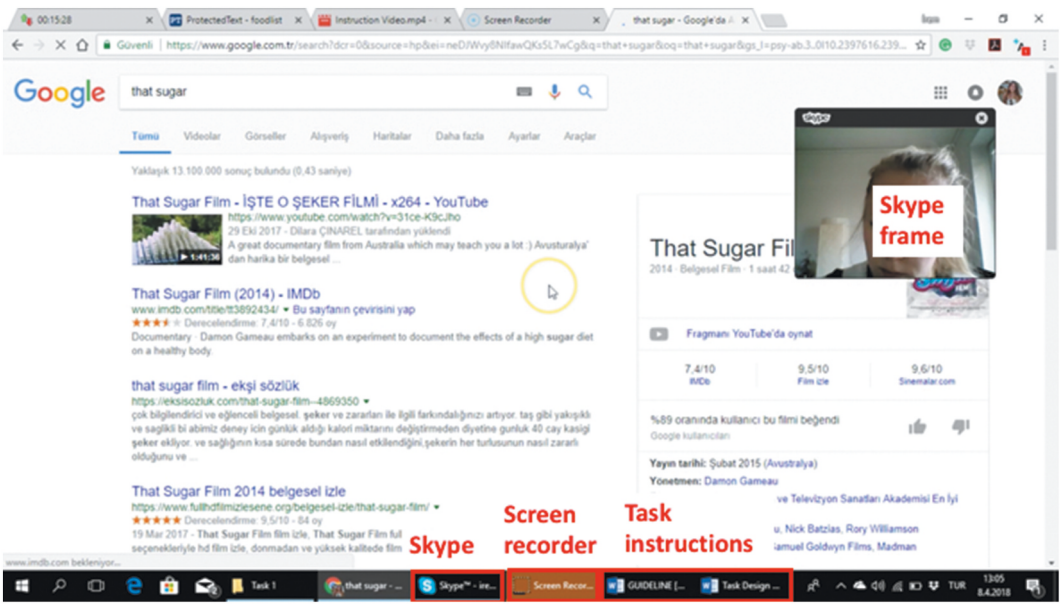

Figure 1. A screenshot of SIN's screen (in bold—e.g., "Skype frame"—are our labels describing items on the screen).

2018), we further restricted our analytic focus in two ways. First, we built a collection of all instances where SIN resorted to a *verbal resource*, such as *I will copy that, wait a minute, I will check*, or *let me check* (n = 29 in DS1; n = 25 in DS2). Second, we excluded all instances where her alert was designed to block off a coparticipant's ongoing turn and action (Ex. 2, line 12), so as to focus on alerts that are prospectively oriented. We then organized the collection chronologically and undertook multimodal sequential analysis of all occurrences. In what follows, we provide analyses of selected extracts that illustrate the observed change in SIN's alerting practices over time, which we complement with quantification of the verbal forms employed by SIN for that precise purpose across the two data sets (see Tables 1 and 2). As it was not possible to identify any relevant cases in the final five weeks of DS1 (mainly due to recording quality) and the last week of DS2 (due to short length of the recording), we present cases over 13 weeks in DS1 and over two weeks in DS2. The collection of cases over time shows a progressive streamlining toward a single "passe-partout" formula for alerting: *let me check*.

Analysis I: Change over 13 weeks

In this section we track change in SIN's participation in the VMIs from their very start over a period of 13 weeks (DS1). In the transcripts, numbers in boldface combined with a hashtag such as 1#, 2#, etc., signal the start of SIN's screen-based activities as described and illustrated in a separate line, and #1, #2, etc., signal the end of such activities (cf. Balaman, 2019). These activities are *not* accessible to coparticipants. Our target action/construction is highlighted in boldface.

Extract 1 occurs only four minutes into SIN's first task-oriented VMI experience in DS1. SIN, NUR, and DEN have to find the word *daguerreotype* (a type of photograph named after its inventor, Daguerre) and for that purpose are given the textual and visual cues that appear on the task interface (see 1#) on their respective screens.

Extract 1. DS1-W1-T4-copy – 04:46–05:00

```
1   NUR:  ↑girls look at the google: (.)
2         for the photo or the for the explaining.
3         (0.3)
4   SIN:  1#oka:y.=
5   DEN:  =oka::y.
6         (1.1#1)
```

1# SIN moves the cursor towards the textual clue area, highlights the clue (Fig. 2), and completes highlighting at **#1**.

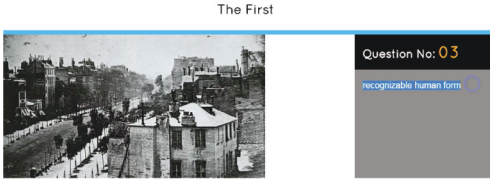

Figure 2

```
7         ((single click sound from SIN's screen))=
8   SIN:  =°i will copy that (.) [and°
9                                [((single click sound from SIN's screen))
10        (2#2.5)
11  DEN:  >you don't need-< (0.8) ↑to copy:
12        (2.0)
13  SIN:  °#2huh-° (1.0) °hmmm°
```

2# SIN pastes copied textual clue to her browser (fig.3), executes a web search, and the results appear at **#2**.

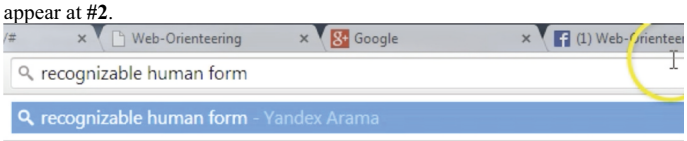

Figure 3

The extract starts with NUR's invitation to her coparticipants to consult their screens. This is acknowledged by both SIN and DEN's *okays* (lines 4/5), with syllable lengthenings which suggest that SIN and DEN may both already be engaged in a search. Simultaneously with her *oka:y*, SIN initiates a screen-based activity (1#)—which is not observable to her coparticipants—and pursues it up to line 6 (#1). She then explicitly, but in low voice, formulates what she is about to do next ("i will copy that," line 8). Through this informing, she both prospectively accounts for her subsequent silence (line 10)—displaying that silence as task-related rather than being due to off-task activity—and bids for the suspension of turn-taking. Although her "and" (line 8) projects a continuation of her turn, she instead engages in the precise screen-based activity (2#) she had foreshadowed. The silence that follows demonstrates that coparticipants treat SIN's alert as what it is: a bid for time to consult her screen while suspending talk-in-interaction.

It is noteworthy how minutely SIN coordinates her talk with her physical manipulation of the screen, as reflected in the movement of her cursor. The initiation of her screen-based activity coincides exactly with the onset of her "oka:y" in line 4, where she moves the cursor to highlight a cue. She completes highlighting in line 6 (see also the click sound from her computer, line 7) exactly before uttering her verbal alert "i will copy that." And she engages in pasting the copied textual clue to her browser to execute a web search exactly after having uttered her verbal alert (2#, line 10; see the click sound, line 9, occurring simultaneously with her "and," line 8). Finally, SIN marks the appearance of search results on her screen (line 13, #2) with "huh-° (1.0) °hmmm°," thereby publicly displaying her continued orientation to the screen. Aspects of SIN's screen-based activity and her talk are inextricably intertwined.

What is specific about this excerpt is not SIN's actual screen-based activity (copy and paste) but the fact that she accounts for her upcoming silence by making the exact nature of her activity explicit ("i will copy that," line 8). This is strikingly different from what we observe SIN do in the last weeks of recording. By identifying the precise activity she is engaging in, SIN's informing lends itself to be countered by coparticipants: In line 11, DEN sanctions SIN's projected activity (">you don't need = < (0.8) =↑to copy:"), possibly because copying is seen as delaying task accomplishment. Yet SIN pursues the copy-and-paste process (2#-#2).

The extract testifies to the fact that, from the very start of the recordings, SIN orchestrates an array of interactional resources to preemptively account for breaks in the progressivity of talk *in sequentially sensitive ways*, including a simple "okay" (understood based on its sequential location in response to an invitation to check the screen), as well as nonlexical vocalizations such as *huh*, which, again due to their location (but see also the lower volume), work as displays of a continuation of an ongoing screen-based activity.

The next extract comes from the fifth week of DS1, illustrating further means that SIN deploys to suspend talk.

Extract 2. DS1-W5-T4-try-wait – 08:05–08:39

```
1   NUR:  >↑google the picture.<
2   SIN:  what?=
3   NUR:  =google the picture SIN.
4         (0.3)
5   SIN:  1#eh ok-
6         eh: >wait a minute<,
7         oka:y,
8   NUR:  a:ndhh (0.8) <you will see [some>
9   ZEH:                             [↑ha:hm
10        (2.4#1)
```

1# SIN changes the tab to image search page with results already listed and holds it still until **#1**.

11 NUR: [2#]key and [peele (.) and something.
12 ZEH: [you see some pict↑ure[#2]s[3#] and the first pictures erm=

2# SIN returns to the task interface and holds it open until **#2**.

13 SIN: =↑wait (.) °wait wait°.
14 ZEH: ↑the answer is written.
15 (.) in the first picture.[#3]

3# SIN returns to the image search results and holds the page still until **#3**.

16 ([4#]5.2)
17 SIN: okay.
18 (.)
19 SIN **i will try >wait a minute<.**
20 (5.0)
21 NUR: °[#4]oka:y°.

4# SIN scrolls down to the part with the answer (Fig.4), returns to the task interface, submits the answer, and the task is completed at **#4**.

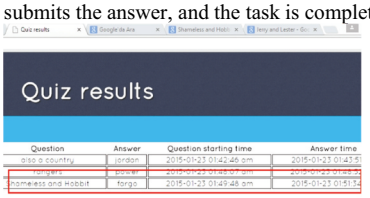

Figure 4

SIN responds to NUR's instruction producing nonlexical vocalizations plus what appears to be a cutoff *okay* (line 5). SIN's verbal and bodily conduct (manipulation of the mouse as seen through cursor movements on the screen) are again closely synchronized: She initiates a screen-based activity (1#) simultaneously with her turn-initial "erm." She then offers an overt request for waiting ("wait a minute," line 6), which is not granted: After SIN's "oka:y" (line 7) with syllable lengthening projecting more to come, NUR continues hinting (lines 8/11), and ZEH starts providing instructions (lines 12/14). Coparticipants' nongranting of waiting time may be due to the fact that SIN's "wait a minute," unlike alerts such as *I will check*, is not unequivocally accountable as related to screen-based activity and, by extension, to task accomplishment. In its context of occurrence, it displays resistance to the instruction just provided by NUR (lines 1–3). This is further illustrated in SIN's subsequent "↑wait (.) °wait wait°" (line 13), again unsuccessfully deployed to cut off ZEH's instruction. SIN's multiple saying of *wait* bears out what Stivers (2004, p. 260) has described in the following terms: "Speakers of multiple sayings communicate their stance that the prior speaker has persisted unnecessarily in the prior course of action and should properly halt course of action." SIN's request hence strongly manifests her treating "private" screen based-activity and "public" talk-in-interaction as practically not deployable simultaneously.

Our focal interest here is on SIN's subsequent alert (line 19), which occurs in a different sequential context than the two preceding uses of *wait*. During a prolonged silence (1ine 16), SIN first scrolls down on her screen to the part of the page with the choice of answers (4#), offers an "okay," and then resorts again to a verbal alert, overtly projecting a screen-based activity ("i will try >wait a minute<," line 19). By contrast to her preceding *wait*-imperatives, this verbal alert offers an accountable display of SIN's engagement in screen-based activity: It identifies (through inference rather than through explicit mention, as in Extract 1) the nature of the screen-based activity—trying to submit the correct answer (#4)—and thereby qualifies her ensuing silence as task-related and ultimately allows her to secure time (line 20) for fully focusing on screen-based activity without being sanctioned nor interrupted by talk. Coparticipants' conduct during the 5-second silence (not noted in the transcript) shows their

understanding and acceptance of SIN's screen orientation: ZEN and DEN return to the task interface and scroll down to the part where they would see whether SIN found the correct answer. NUR continues gazing at her screen.

This second extract provides further examples of the various interactional resources deployed by SIN for the purpose of signaling incipient or continued screen-based activity. Although her "↑wait (.) °wait wait°" (line 13) is deployed in a responsive way to block off coparticipants' talk (and is hence not part of our collection), her "eh: wait a minute" (line 6) and "i will try >wait a minute<" (line 19) are designed to prospectively secure space for SIN to orient to her screen—yet they are responded to differently by coparticipants. Only the latter is functional in preemptively accounting for incipient silences as being task-related: It secures further time for a screen-based activity.

Taken together, Extracts 1 and 2 show SIN's distinctive management of the two concurrent courses of activity—screen search and talk-in-interaction: While treating engagement in one or the other as mutually exclusive, she often—yet at this point not consistently—orients to the relevance of accounting for breaks in progressivity *as being task-related* and, by virtue of that fact, manages to suspend the progression of talk-in-interaction in a mutually agreed-upon way while by the same token advancing task accomplishment. This becomes more systematic over time. As shown in the following excerpts, from week 8 on, coparticipants consistently display understanding of SIN's bids for suspension of talk as being instrumental in the moving forward of the joint goal-oriented activity, allowing SIN to focus exclusively on retrieving task-relevant information from the screen.

Extract 3 shows one of the first occurrences of an alert comprising the verb *check*. SIN's alert is again formatted as an informing about her future action and secures a 21.2-second suspension of talk. Participants are expected to identify a particular tweet and submit the author's name as the correct answer.

Extract 3. DS1-W9-T4-i'll check – 02:16–02:55

```
1   DEN:  1#uh- i said eh: n- nineteenth of february there is a
2         (0.4) a::: retweet.
3         (1.3)
4         and there is also=
5   SIN:  =↑okay.=
6   DEN:  =a-
7   SIN:  okay uh i'll check okay,
8         (21.2)
9         thursday afternoon#1
```

1# SIN holds the cursor still (Fig.5) until the target page loads at #1.

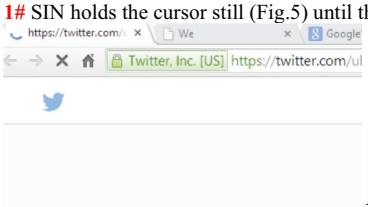

Figure 5

```
10        ((clears her throat))
11        (1.6)
12        thur- thursday (.) what was the date?
13        (0.7)
14  DEN:  [nineteenth of
15  SIN:  [thursda:::y
16  DEN:  ↑o:h not february (.) sorry.
```

SIN is engaged in screen search from the very start of the extract (1#). DEN's turn (lines 1–4) is blocked by SIN's "okay" (line 5), and SIN then alerts to her screen-based activity ("okay uh i'll check

okay"), the nature of which is here inferable from her turn's occurring in response to NUR's preceding turn. Coparticipants respond by granting suspension of talk but also by engaging in their own screen-based activities, testifying to joint engagement in task accomplishment: DEN enters candidate answers on the task interface while ZEH remains on the Twitter page (not noted in transcript).

Extract 3 comes from week 9 of the recordings, and it is from this time on that we see SIN start to build her alerts near to exclusively around the verb *check*. Progressively, "I will check (X)" becomes the standard action format she uses, with *X* being either expressed as a grammatical complement or remaining implicit but inferable from context.

Extract 4 starts with a 4-second silence, before which coparticipants were engaged in an off-task conversation. SIN's *well*-prefacing of her turn (line 2) marks the reestablishment of her orientation to the task, which consists in finding a book title (*Survivors*) that bears the same name as a famous TV show (*Survivor*).

Extract 4. DS1-W12-T4-wait-I will check – 01:10:20–01:10:53

```
1          (4.0)
2   SIN:   1#(well) i will check survivor again.
3          (1.7)
4          #1and go to tv sekiz (.) $websitehh$
```

1# SIN searches for survivor; the results are displayed at **#1** (Fig.6).

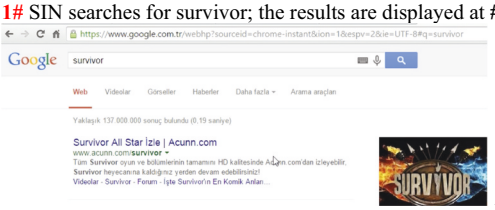

Figure 6

```
5          (1.5)
6   NUR:   >okayh,<
7          (0.7)
8          wh- (0.5) which celebrities (0.5) th- (0.2) are on survivor?
9          (2.0)
10  SIN:   2#ahm↑mmm↓ (1.0) oka:y,
11         (0.9)
12  NUR:   pascal ↑nou↓ma
13  SIN:   #2wait.
```

2# SIN moves the cursor on top of the first result, clicks on it, and a new page loads.

```
14         (1.5)
15         3#yes:
16         (2.1)
17         [a:nd
18  NUR:   [uhm
19         (3.1)
20  SIN:   wait I'll check it.
21         (0.8)
22  NUR:   okay.
23         (2.7)
24  SIN:   >#3humm.<
```

3# The browser loads the new page (Fig.7), SIN scrolls down to the part where survivor contestants are listed and holds the cursor still next to the list at **#3**.

Figure 7

The two tokens of *I will check*, just like the alerts shown in the preceding extracts, occur in an environment of prolonged silences and prospectively secure space for SIN's screen-based activity while suspending talk-in-interaction, i.e., they are treated by coparticipants as bids for SIN to fully focus on screen search. They are again offered as informings about SIN's incipient doing (future tense), the second of them being combined with the imperative form *wait* (similar to Ex. 2). And they are rather transparent as to the very nature of SIN's screen-based activity. ("I'll check it," line 20, can be heard as tying back to NUR's questions as to which celebrities are on *Survivor* [line 8] and is attended to by NUR as a call for suspension of talk: NUR waits while gazing at her screen). Furthermore, the extract provides an example of how extended such screen-based activities can be and how efficiently SIN periodically intersperses these with nonverbal vocalizations and minimal signs of her being engaged in a search (lines 8–20), after which coparticipants refrain from taking the floor (and if not, are sanctioned by SIN, lines 12/13). It is only in line 24 (#3) that SIN finally reaches the page where *Survivor* (the TV show) contestants are listed, which she receipts with a short ">humm.<."

In sum, from her very first participation in task-oriented VMI, we see SIN deploy a range of practices for the purpose of preemptively accounting for breaks in the progressivity of talk-in-interaction by alerting recipients to her own incipient or ongoing screen-based activity, thereby issuing a bid for suspension of talk. Focusing on those practices that have a verbal component and are prospectively oriented, we observe both continuity and change over the first 13 weeks of SIN's VMIs. In terms of continuity, SIN's alerts are oriented to as such and aligned to, as coparticipants usually treat talk-in-interaction as momentarily suspended. They are most often proffered as informings, in declarative form and future tense, that make aspects of the nature of the search activity either explicit or easily inferable. By virtue of their being displayed as search-related, and hence task-related, incipient silences are accounted for and recognized as instrumental in moving the task accomplishment forward: The progressivity of the overall joint goal-oriented activity takes precedence over the progressivity of talk-in-interaction. By contrast, when silences or screen activities are not seen as fostering task accomplishment, they lend themselves to being sanctioned by coparticipants (see the "copy" in Ex. 1 and the "wait" in Ex. 2). The way SIN's alerts are attended to by coparticipants manifests *joint* orientation to the normative organization of the multiactivity at hand: Participants orient to the "exclusive orders" of the concurrent courses of activity, to their related rights to bid for suspension of talk-in-interaction for task-related purposes and their obligations to grant such suspensions, and they treat the navigation between activities as an integral part of the cooperative task accomplishment.

In terms of change, exceptions to the previous are found only in a few instances between weeks 1 and 5 (Ex. 1, 2), where SIN's bids for suspension of talk are occasionally sanctioned or not granted. Yet what changes are the precise verbal resources SIN uses for her alerts. This is schematically illustrated in Table 1.

Table 1 indicates both the pathway of change and the sudden "explosion" (relative frequency) of the *I'll check* construction. SIN starts off with a variety of morphosyntactic and lexical formats (*I will copy;*

Table 1. Overview of the collection in DS1.

Week 1	**I will** ***copy*** that (Extract 1)
Week 4	***wait*** a minute
Week 5	***wait*** a minute (Extract 2) **I will** ***try*** >***wait*** a minute< (Extract 2)
Week 8	**I will** ***try***
Week 9	**I'll** ***check*** okay (Extract 3)
Week 12	**I will** ***check*** this clue **I'll** ***check*** oka:y here **I will** ***check*** tv **I will** ***check*** the other **I will** ***check*** survivor again (Extract 4) ***wait*** **I'll** ***check*** it (Extract 4) ah **I will** ***check*** survivor's u k version again
Week 13	**I'll** ***check*** (.) windows (.) office **I will** ***check***

wait a minute; I will try) used up to week 8 and then streamlines her alerting practices based on the recurrent fixed expression *I will check/I'll check*, which makes up 100% of all the occurrences from week 9 on. Interesting is the fact that SIN—an L2 speaker—*does* have the verb *check* in her repertoire before week 9[1] but does not (yet) use it as part of the practice and within the action context where it will later become the standard. Now let us turn to what we see SIN do 4 years later, in her second round of participation in task-oriented VMIs.

Analysis II: 4 years later

Although the VMIs in DS2 consist of collective guessing activities comparable to DS1, here SIN interacts with one single coparticipant. By now, the verb *check* has become the prototypical verb island (Tomasello, 2003) around which SIN constructs her alerts; these alerts continue to occasionally shape up as the precise multiword expression *I'll check*, yet another expression—*let me check*—now is the predominant resource used by SIN for alerting coparticipants to her incipient screen-based activity.

Extract 5 is taken from a movie guessing task. It starts while SIN is searching for cues on her screen and typing the movie title "That Sugar" into the search line. Just like in DS1, SIN's alerts (lines 5, 11, 33) occur in environments of prolonged silences during which she is engaged in manipulating items on the screen—an activity that is, again, not inspectable to her coparticipants. Different from the preceding data set, however, the lexicosyntactic constituency of her alerts is *let me check* and a cutoff *let me-*.

Extract 5. DS2–That Sugar – 19:05–20:01

1 SIN: uhhmmmm
2 (3.8)
3 errmmmm
4 (3.1)
5 °let me° (0.7[1#]) check (0.2)
6 /I/ /em/ /di/ /bi:/,
7 (2.9)
8 >i think it[#1]'s[2#]< (0.3) er:m ([#2]0.7) [3#]erm:

1# SIN moves the cursor up and changes the tab to the IMDB page - **#1**.
2# SIN scrolls down to the guideline doc. minimized on the task bar (Fig.8) and opens it at **#2**.

Figure 8

9 [#3]sashhh (.) am I allowed to °say that?°

3# SIN moves the cursor on to the paragraph on the Word document which includes information about task rules.

10 ([4#]0.6)
11 just- >°let° me-<
12 (2.2)
13 OKAY[#4].

4# SIN scrolls up on Word; OKAY is produced when the cursor is on the "things to say" list available in the Word document.

[1]The first occurrence is in an invitation to the group to proceed in a given way: "We should check the post" (DS1-W8_T4).

14 (0.5)
15 PAT: yeahh [5#just
16 SIN: [errm
17 PAT: $sa:: ↑hah hah hah$
18 SIN: it's (0.7) originated i:n Australia,
19 PAT: °australia [okay°
20 SIN: [erm
21 (1.0)
22 and
23 (2.2)
24 the (0.3)
25 <director of the movie also the sta:r (0.3)
26 of the movie.#5>

5# SIN returns to the IMDB page, scrolls up and down, holds the cursor still on the part where the director information is provided **#5**.

27 (6#2.1)
28 PAT: oka:y.
29 SIN: and ↑in the movie
30 (1.9)

31 er:mmm
32 (2.8)
33 °#6**let me che:ck,°**

6# SIN scrolls down to Word, opens the guideline document, and moves the cursor closer to the "things to say" part (Fig.9 schematizes the mouse movements from #3 through 6#).

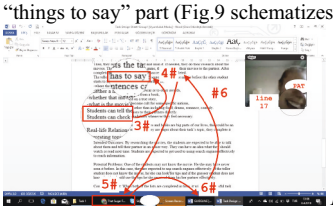

Figure 9

34 (2.1)
35 uhmmmmmm
36 (0.6)

At the start of the excerpt, SIN's nonlexical vocalizations suggest that she is engaged in consulting her screen. In line 5, she offers an explicit alert about a precise incipient screen-based activity (checking the IMDB page), which she engages in concurrently with uttering the alert (#1). Importantly, and different from the earlier recordings, SIN's alert is not produced as an informing about a future course of action; rather, she resorts to an imperatively formatted construction that has been treated as a conventionalized indirect request seeking permission (West, 1990), akin to a suggestion (Aijmer, 1996): *let me X*. As Hoey (2020) has demonstrated, with *let me X*, the speaker only apparently solicits the recipient's consent, but in fact leaves little latitude for resistance: The speaker presumes the recipient's permission. Hoey shows that often in interaction, the construction does not function as a request but as a means of "self-authorizing" one's own action—an action that is either already under way or incipient. This is exactly what we observe in the data, and it is by virtue of this "self-authorization" that the *let me check* request format converges with that of the informing format discussed previously. Both project the speaker's incipient action—an action that is treated by all participants as necessitating the suspension of talk.

The extract demonstrates how SIN resorts to verbal alerts in alternation with nonlexical vocalizations (lines 1, 3, 31, 35) as well as grammatical projection (she suspends syntactic trajectories at points of "maximum grammatical control," Schegloff, 1996; lines 22/24, 29) to make her coparticipant

understand that she is searching. Her verbal alerts alternate with these other means in nonrandom ways; the former specifically prolongs a search alert where talk has been suspended for quite a while (e.g., lines 4/5, 32/33). By remaining silent (lines 7, 12, 34), PAT in turn displays recognition and acceptance of SIN's "let me (check)" as indicating her orientation to the screen, which PAT treats as a valid reason for suspending talk. So SIN's use of the construction appears to entail exactly the same interactional consequences as her varied uses of informings documented in DS1.

Further noteworthy is again how closely *let me check* is coordinated with SIN's on-screen activity. Starting in line 5, "°let me° (0.7) check (0.2) /I//em//di//bi:/" coincides with SIN's scrolling up, and "check" is delivered simultaneously with SIN's clicking on the tab of the IMDB page (1#). Such synchronization is achieved through the speed of delivery of the verbal segment, with its interspersed 0.7-second pause. This is also observable in SIN's subsequent offering of the fragment "let me-" (line 11), again ensuing in a longer silence (line 12). As shown in #4, SIN scrolls up the text to a precise spot in the document, which is synchronously marked in talk with the loud production of "OKAY" (line 13), possibly as a way of making the end of her preceding "private" search public. Finally, in line 33, SIN's "let" coincides exactly with holding the cursor on the "things to say" part of the guideline document (#6).

This extract is representative for DS2: SIN predominantly uses the lexically specific construction *let me check* for the purpose of alerting her coparticipant to her own screen orientation and thereby making a bid for suspension of talk. Importantly, as opposed to DS1, in DS2 precise identification of the projected action (explicit or by inference) is less frequent, and often the construction does nothing more than project the speaker's incipient orientation to the screen. Consider Extract 6 (discussed in more detail in Balaman & Pekarek Doehler, in press), where the nature of the actual screen activity remains indeterminate:

Extract 6. DS2_Woman on Top–30:36–31:19

```
1   SIN:  yeah forks over knives
2   PAT:  yes exactly (0.4) nice
3         (3.2)
4   SIN:  hmmm
5         (1#1.8#1)
          1# SIN opens up the browser.

6         2#let me che:ck#2
          2# SIN moves the cursor up to the address bar and highlights the existing entry to remove.

7         (3#4.9#3)
          3# SIN types "Woman on Top" to the address bar; presses a keyboard button to execute a Google
          search, and search results appear on #3.

8         the movie
9         (1.2)
10        erm::: (.) the year of the movie i:s
11        (0.7)
12        two thousand
```

Significant in regard to these underspecifications is that the occurrences of just the fragment *let me-* (Ex. 5, line 11) and of *let me check* without object (Ex. 5, 6) are still oriented to by coparticipants as bids by SIN for time to consult her screen and suspend talk (see the ensuing lengthy silences).

Although we have here specifically focused on SIN because of the availability of longitudinal data for her, the practices and resources found in DS2 are not exclusive to her: We see other participants in the data recurrently deploy coordination work similar to SIN, including the frequent use of *let me X* (see Balaman & Pekarek Doehler, in press, for an example of *let me see*).

Table 2. Overview of the collection in DS2.

Week 1	***I will check*** it o:n Google
	let me just make a search
	let me (0.7) check (0.2) imdb (Extract 5)
	just let me (Extract 5)
	let me che:ck (Extract 5)
	let me just check it
	let me check
	let me che:ck (Extract 6)
	let me just check the video for a second okay
	I will check
	just let me
	let me just check
Week 2	**let me check** lint
	let me check
	I'll check the answer
	let me check
	I will check it (.) wait
	let me check
	let me check its English
	let me write okay
	so *I will* write the baklava
	okay ***I'll check***
	let me check the calendar

The collection of cases from DS2, as schematically presented in Table 2, shows that SIN now publicizes her private incipient screen-based activity mainly through the deployment of a highly routinized resource: *let me check*.

Compared to four years earlier, SIN's alerting practices now build on a restricted set of constructions: The lexically fixed multiword expression *I will/I'll check* is still used (22%, $n = 5$), but 74% of her alerts (17 out of 23) in DS2 are based on the constructional schema (let + me + [X])—a conventionalized indirect request format—rather than being implemented through informings (*I'll check*). The lexically specific format *let me check* represents more than half (56%, $n = 13$) of all verbal alerts issued by SIN: It has become the typical expression used for signaling her screen-based activity and prospectively accounting for breaks in the progressivity of talk as being due to that activity. By the same token, the often precise (explicit or inferable) identification of the nature of SIN's projected on-screen activity observed in DS1 now cedes its place to alerts that most often remain unspecified as to that nature (Ex. 5, 6). Significant in this regard is that eight out of 13 occurrences of *let me check* do not have an object complement, and two further occurrences are mere *let me* fragments, yet are perfectly functional interactionally (Ex. 5).

Discussion: Recalibrating resources for action when adapting to new situations, languages, and/or media

In the data under scrutiny, we observe participants attend both to the specific constraints and affordances of VMI (Licoppe & Morel, 2014) and to generic organizational principles of social interaction (Schegloff, 2007). Orientation to "procedural consequentiality of [video] mediation" (Arminen, Licoppe & Spagnolli, 2016, p. 305) materializes through how participants work to articulate screen-based activity and talk-in-interaction in ways that are appurtenant to their unfolding task-oriented interaction; they make relevant an "exclusive order" (Mondada, 2014) of the multiactivity at hand, and they deploy methodic means to switch between activities (interacting with others, manipulating their own screens) in accountable ways. Participants' normative orientation to fundamental principles of social interaction in turn is reflected in their continuous concern with accounting for breaks in the progressivity of talk-in-interaction and for deviations from the principle of "no gap" as being task related, i.e., due to screen-based activities deployed in view of joint task accomplishment.

Taken together, the aforementioned highlights some of the eminently situation- and medium-specific exigencies for social coordination entailed by the highly specialized activity context under scrutiny.

Our analytic focus was on the methodic ways in which one participant, SIN—initially a novice—dealt with some of these exigencies across her task-oriented VMI history. We examined SIN's verbal practices to alert coparticipants to her own incipient screen-based activity, thereby offering a bid for halting the turn-taking machinery while prospectively accounting for the suspension of talk-in-interaction as being task related. Our exclusive focus on vocal resources was motivated by the fact that bodily-visual altering resources were not observably deployed in the sequential environment studied, most likely due to highly limited visual accessibility of participants to each other. Although our target participant is not only a novice to VMI but also a second-language speaker, we see the observed change in her practices as pertaining not simply to second-language acquisition (note that she used the verb *check* already in the first data set, but in different environments and for other purposes); rather, we understand that change as part of the larger process of how people recalibrate their resources for action when adapting to new situations, languages, and/or media.

Three dimensions of change over time were identified:

(1) Initially constructing her alerting practices based on a variety of verbs *(wait, try, copy, check*) and two syntactic patterns (imperative and declarative), over an initial period of 13 weeks our novice participant's coordinative practices routinized into the recurrent use of a verbal alert constructed around the verb *to check* (*I will check*); four years later, that routinization eventually further evolved into the use of the construction *let me (X)*, most often lexically specific *let me check*, as a typical action format for the purpose described here.
(2) The precise nature of the screen-based activity so projected became less explicit or inferable and ultimately remained often unspecified.
(3) The actions used as vehicles for alerting (see Schegloff, 2007 on the "double-barreled" nature of action) changed from informings about the speaker's incipient doing through first-person future-tense declarative formats (e.g., *I will check*) to the predominant use of what has been discussed as a prototypical (indirect) request format—the imperative form *let me (X)*.

The latter point begs an important question: Does this change in vehicular action materialize a shift in mutual entitlements, the informings embodying the speaker's absolute entitlement to perform the projected action, and the *let me (X)*, by contrast, conferring entitlement to the recipient to grant (or not) the request for action? We would like to argue that this is not the case. Rather, the two vehicular actions work in related ways to project an activity (screen-based search) that is treated by participants as suspending another activity (talk-in-interaction).

For one thing, as Hoey (2020) recently documented, *let me (X)* is distinct from other imperative formats (such as *wait*, Ex. 2) in that it allocates more agency to the speaker and often functions as a means of self-authorizing action: "while it treats the recipient as able to resist [. . .], it does not provide an express location for them to do so" (p. 31). This is clearly observable in the data: Just like with her informings, SIN starts to deploy the projected activity often simultaneously with the delivery of *let me check*, thereby eschewing any sequential opportunity for coparticipants' granting. She hence displays a "presumption of permission" (Hoey, 2020, p. 6) that can be explained by the fact that the projected activity is understood by all participants as instrumental to the collaborative task at hand.

For another thing, both the informings and the *let me check* are attended to in related ways by coparticipants. While prima facie informings, just like requests, make recipient response relevant (Thompson et al., 2015), such response is absent in the data; thereby, coparticipants tacitly display understanding of SIN's alerts as bidding for suspension of talk-in-interaction. This testifies to the fact that coparticipants orient to SIN's alerts not as informings nor as requests but as devices that project a switch between two exclusive orders of activities and by the same token account for the need to suspend talk as being task related.

Yet formally, the two differ: SIN's informings are proffered as speaker-centered (egocentric) formulations (*I will*), while her *let me (X)* materialize as intersubjective formulations. As grammatical imperatives, the latter formally embody a claim of entitlement to issue a directive and by the same token display strong deontic authority on the part of the speaker (Stevanovic, 2018). At the same time, however, they formally construct the recipient as a coagent, as the "granter": There is a sense of an appeal to the recipient—but only on the surface. Now this switch from a more egocentric (*I will*) to a more intersubjective (*let me*) format, while it does not entail distinct interactional consequences, may itself be symptomatic for SIN's increasingly embodying the *collaborative* nature of task and orientation to the interaction as driven by a *joint* goal.

Importantly, the shift from informing to *let me check* implies a decrease in the transparency of the projected activity. This is partially entailed by the very nature of the verb *check*, which, in its contexts of occurrence, does nothing more than indicate the speaker's consulting (checking) her screen. Such indeterminacy of the projected screen activity is also reflected in the change in constituency of the constructions used, the *let me check* being often used in the intransitive form, i.e., without an object complement, and twice occurring just as *let me-*. Such syntactic-semantic reduction is both a sign of the routinization of the construction and a functional aspect of what it is designed to do in interaction: By means of *let me (check)*, the speaker does nothing more than alert to an unspecified screen-based activity—but neither does she do less than that (see, e.g., her "wait" in Ex. 1). This very lack of specification precludes the possibility of the projected action to be subject to refutation by coparticipants (as was the case with the action-identifying "I'll copy that," Ex. 1) and at the same time allows for the generalized "passe-partout" use of the construction: It provides an eminent case for vagueness as a member's resource.[2] By the same token, the *let me check* displays SIN's experientially built understanding of the sufficiency for her to issue an alert to her screen-based activity without requiring identification of the type of activity she is about to undertake. In this sense, her alerts in DS2 are reflexively tied to an experientially accumulated shared understanding of the exclusive order of activities at hand and the related rights to suspend talk for task-related purposes. Through its intersubjective nature as well as its underspecification, the *let me (check)* is specifically "fitted" to do the job it is designed to do in its precise interactional environment: It routinizes as a "social action format"—"a grammatical format for sequentially- [and situationally]-specific actions" (Fox, 2007, p. 204)—i.e., as an operational solution to a practical problem of social coordination within the multimodal ecology of "video-in-interaction" (Licoppe & Morel, 2012).

This result, however, poses a certain puzzle, as it traces a path from lexical and constructional diversity to a single lexically specific construction used as an interactional resource. This contrasts with evidence of language development as a usage-driven process involving increased schematicity of language knowledge emerging from use (Ellis, 2012; Tomasello, 2003). It also stands in contrast to existing CA-SLA findings on the development of L2 grammar as an interactional resource (see the section on "Background" above), which document a diversification of the action contexts where a given construction occurs.

How can we explain such a contrast? The fact is that developmental studies have overwhelmingly proceeded by tracking lexicosyntactic resources over time; they have *not* started from actions. Yet when we start from actions, and track the resources deployed therein, we begin to see the other side of the coin: Although the diversification of grammar may allow people to accomplish more and more diverse actions in more and more varied situations in increasingly context-sensitive ways (Pekarek Doehler, 2018; Pekarek Doehler & Berger, 2018), recurrent dealings with precise actions within precise situations and precise local contingencies may have the effect of stabilizing and ultimately routinizing the grammatical solutions deployed therein (Pekarek Doehler, in press). The diverse repertoire of grammatical resources that people develop throughout their lives may be nothing more than the product of people's repeated dealing with the myriad of social-interactional exigencies they encounter across their interactional histories (see also Deppermann & Schmidt, 2021/this issue, for a longitudinal study of participants' appropriating the meaning of an expression). Although supporting the idea that

[2]We thank Arnulf Deppermann for having drawn our attention to this point.

linguistic structure arises from the contingencies of use and specifically from talk-in-interaction (Bybee & Hopper, 2001, p. 3; Fox, 2007, p. 299; Hopper, 2011), this foregrounds the eminently activity-bound nature of grammatical resources and interactional competence and their development. The findings ultimately stress that interactional competence is not about having more forms but about deploying contingent practical solutions for getting locally relevant interactional work done.

The findings boil down to identifying a progressive simplification of a social action format in the specialized context at hand. They reverberate with Deppermann's (2018) evidence on how a driving instructor over time progressively simplifies the action formats of his instructions, both grammatically and in terms of their informational contents, as part of how he adapts to the expertise and shared knowledge of his student. Equally, our results show how iterative engagement in the same joint activity (or task) and iterative accomplishment of a given action therein inform the design of that action over time and drive the development of a grammatical usage pattern for accomplishing and coordinating actions in locally contingent ways. In this sense, grammatical practices behave just like other practices: Repeated actions may, over time, become reduced in structure (e.g., Clark & Marshall, 1981; see here the intransitive use of *check*), and such simplification is—just like diversification—an integral part of people's developing shared methods for action as part of their moving from being "not yet competent members" (Schegloff, 1989) to becoming "competent members."

Conclusion: The emergence of grammatical patterns as practical solutions to recurrent social-interactional exigencies

The aforementioned has profound implications for how we understand grammar-for-interaction, with possible reverberations on how we see the evolution of grammar out of interaction. The current analysis, though limited to one participant's practices for social coordination for a precise purpose and in a precise interactional setting, offers a glimpse on how grammatical usage patterns may be the cumulative products of people's interactional histories.

Structural patterns are accumulated through repeated engagement in interaction over time. These patterns show a certain degree of stability not due to abstract combinations rules but as epiphenomena of frequent combination of use-in-interaction, i.e., as adaptations to recurrent usage problems that have become conventionalized through repetition (Bybee & Hopper, 2001; Fox, 2007; Hopper, 2011). In this sense, they may best be understood as the *aggregates of repeated verbal solutions to interactional exigencies*—solutions that are experientially ratified as locally valid and that, through iterative processes of such experiential validation, ultimately routinize as functional responses to recurrent interactional needs (Pekarek Doehler, in press). They are locally accountable as such and therefore may become shared solutions, mutually recognizable solutions, and in the end "conventionalized" solutions for people's dealing with recurrent practical problems.

The need for projecting incipient courses of action—enabling anticipation by coparticipants—is one such central interactional problem. The constructional pattern that we saw routinize over time was tied to the fundamental need of projection, i.e., alerting coparticipants to what comes next, and its correlate, anticipation. Deppermann and Streeck's (2018, p. 6) observation regarding recurrent projection practices across language communities may as well apply to the emergence of grammatical patterns for social interaction: "As adaptations to the needs for projection and anticipation, they favor re-use of prior solutions, that is, trust in the iterability of solutions for coordination problems that have previously worked under similar circumstances." Our data opened a window onto how grammatical patterns working as "solutions for coordination problems" under similar circumstances may progressively, through reiteration along interactional histories, become selectively streamlined as the most efficacious and mutually accountable solutions for coping with the local organizational needs for conducting social interaction and maintaining intersubjectivity. This is part of how people recalibrate their interactional competences when faced with novel situations, involving novel interactional ecologies. In this understanding, sedimented patterns of action (practices) and sedimented patterns of grammar (grammatical constructions) may have the very same origins: people's dealing with the

local contingencies of social interaction.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Simona Pekarek Doehler http://orcid.org/0000-0002-3592-5076
Ufuk Balaman http://orcid.org/0000-0001-5092-9414

References

Aijmer, K. (1996). *Conversational routines in English: Convention and creativity*. Longman.
Arminen, I., Licoppe, C., & Spagnolli, A. (2016). Respecifying mediated interaction. *Research on Language and Social Interaction*, *49*(4), 290–309. https://doi.org/10.1080/08351813.2016.1234614
Balaman, U. (2018). Task-induced development of hinting behaviors in online L2 task-oriented interaction. *Language Learning & Technology*, *22*(2), 95–115. https://doi.org/10125/44640
Balaman, U. (2019). Sequential organization of hinting in online task-oriented L2 interaction. *Text & Talk*, *39*(4), 511–534. https://doi.org/10.1515/text-2019-2038
Balaman, U., & Pekarek Doehler, S. (in press). Navigating the complex social ecology of screen-based activity in video-mediated interaction. *Pragmatics*.
Balaman, U., & Sert, O. (2017a). Development of L2 interactional resources for online collaborative task accomplishment. *Computer Assisted Language Learning*, *30*(7), 601–630. https://doi.org/10.1080/09588221.2017.1334667
Balaman, U., & Sert, O. (2017b). The coordination of online L2 interaction and orientations to task interface for epistemic progression. *Journal of Pragmatics*, *115*, 115–129. https://doi.org/10.1016/j.pragma.2017.01.015
Bybee, J., & Hopper, P. J. (2001). *Frequency and the emergence of linguistic structure*. John Benjamins.
Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). CUP.
Couper-Kuhlen, E., & Selting, M. (2018). *Interactional linguistics*. Cambridge University Press.
Depperman, A., & Pekarek Doehler, S. (2021/this issue). Longitudinal conversation analysis - introduction to the special issue. *Research on Language and Social Interaction*, *54*(2), 127–141. https://doi.org/10.1080/08351813.2021.1899707
Deppermann, A. (2018). Changes in turn-design over interactional histories: The case of instructions in driving school lessons. In A. Deppermann & J. Streeck (Eds.), *Time in embodied interaction: Synchronicity and sequentiality of multimodal resources* (pp. 293–324). John Benjamins.
Deppermann, A., & Schmidt, A. (2021/this issue). How shared meanings and uses emerge over an interactional history: *Wabi sabi* in a series of theatre rehearsals. *Research on Langauge and Social Interaction*, *54*(2), 203–224. https://doi.org/10.1080/08351813.2021.1899714
Deppermann, A., & Streeck, J. (Eds.). (2018). *Time in embodied interaction: Synchronicity and sequentiality of multimodal resources*. John Benjamins.
Ellis, N. (2012). Frequency-based accounts of second language acquisition. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 193–210). Routledge.
Eskildsen, S. W. (2011). The L2 inventory in action: Conversation analysis and usage-based linguistics in SLA. In G. Pallotti & J. Wagner (Eds.), *L2 learning as social practice: Conversation analytic perspectives* (pp. 327–364). National Foreign Language Resource Center.
Fox, B. (2007). Principles shaping grammatical practices: An exploration. *Discourse Studies*, *9*(3), 299–318. https://doi.org/10.1177/1461445607076201
Garfinkel, H. (1967). *Studies in ethnomethodology*. Polity.
Goodwin, C. (2018). *Co-operative action*. Cambridge University Press.
Haddington, P., Keisanen, T., Mondada, L., & Neville, M. (Eds.). (2014). *Multiactivity in social interaction: Beyond multitasking*. John Benamins.
Heath, C., & Luff, P. (1993). Disembodied conduct: Interactional asymmetries in video-mediated communication. In G. Button (Ed.), *Technology in working order: Studies of work, interaction, and technology* (pp. 35–54). Routledge.
Heath, C., & Luff, P. (eds.). (2000). *Technology in action*. Cambridge University Press.
Heritage, J. (2007). Intersubjectivity and progressivity in person (and place) reference. In N. J. Enfield & T. Stivers (Eds.), *Person reference in interaction* (pp. 255–280). Cambridge University Press.
Hjulstad, J. (2016). Practices of organizing built space in videoconference-mediated interactions. *Research on Language and Social Interaction*, *49*(4), 325–341. https://doi.org/10.1080/08351813.2016.1199087

Hoey, E. (2020). Self-authorizing action: On *let me X* in English social interaction. *Language in Society*, 1–24. https://doi.org/10.1017/S0047404520000779

Hopper, P. J. (2011). Emergent grammar and temporality in interactional linguistics. In P. Auer & S. Pfänder (Eds.), *Constructions: Emerging and emergent* (pp. 22–44). Walter de Gruyter.

Licoppe, C., & Morel, J. (2012). Video-in-interaction: "Talking heads" and the multimodal organization of mobile and Skype video calls. *Research on Language and Social Interaction*, *45*(4), 399–429. https://doi.org/10.1080/08351813.2012.724996

Licoppe, C., & Morel, J. (2014). Mundane video directors in interaction: Showing one's environment in Skype and mobile video calls. In M. Broth, E. Laurier, & L. Mondada (Eds.), *Studies of video practices* (pp. 143–168). Routledge.

Luff, P., Heath, C., Kuzuoka, H., Hindmarsh, J., Yamazaki, K., & Oyama, S. (2003). Fractured ecologies: Creating environments for collaboration. *Human Computer Interaction*, *18*(1–2), 51–84. https://doi.org/10.1207/S15327051HCI1812_3

Luff, P., Heath, C., Yamashita, N., Kuzuoka, H., & Jirotka, M. (2016). Embedded reference: Translocating gestures in video-mediated interaction. *Research on Language and Social Interaction*, *49*(4), 342–361. https://doi.org/10.1080/08351813.2016.1199088

Mondada, L. (2014). The temporal orders of multiactivity. Operating and demonstrating in the surgical theatre. In P. Haddington, T. Keisanen, L. Mondada, & M. Neville (Eds.), *Multiactivity in social interaction: Beyond multitasking* (pp. 33–75). John Benjamins.

Näslund, S. (2016). Tacit tango: The social framework of screen-focused silence in institutional telephone calls. *Journal of Pragmatics*, *91*, 60–79. https://doi.org/10.1016/j.pragma.2015.10.008

Ochs, E., Schegloff, E. A., & Thompson, S. A. (eds.). (1996). *Interaction and grammar*. Cambridge University Press.

Pekarek Doehler, S. (2018). Elaborations on L2 interactional competence: The development of L2 grammar-for-interaction. *Classroom Discourse*, *9*(1), 3–24. https://doi.org/10.1080/19463014.2018.1437759

Pekarek Doehler, S. (in press). How grammar grows out of social interaction: From multi-unit to single unit question. *Open Linguistics*.

Pekarek Doehler, S., & Berger, E. (2018). L2 interactional competence as increased ability for context-sensitive conduct: A longitudinal study of story-openings. *Applied Linguistics*, *39*(4), 555–578. https://doi.org/10.1093/applin/amw021

Pekarek Doehler, S., & Pochon-Berger, E. (2015). The development of L2 interactional competence: Evidence from turn-taking organization, sequence organization, repair organization and preference organization. In T. Cadierno & S. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 233–268). Mouton De Gruyter.

Pekarek Doehler, S., Wagner, J., & González-Martínez, E. (eds.). (2018). *Longitudinal studies on the organization of social interaction*. Palgrave Macmillan.

Rintel, S. (2013). Video calling in long-distance relationships: The opportunistic use of audio/video distortions as a relational resource. *The Electronic Journal of Communication/La Revue Electronic de Communication*, *23*(1&2), 1.

Sacks, H., & Schegloff, E. A. (1979). Two preferences in the organization of reference to persons in conversation and their interaction. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 15–21). Irvington Publishers.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696–735. https://doi.org/10.1353/lan.1974.0010

Schegloff, E. A. (1989). Reflections on language, development, and the interactional character of talk-in-interaction. In M. H. Bornstein & J. S. Bruner (Eds.), *Interaction in human development* (pp. 139–153). Lawrence Erlbaum.

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 University Press.

Schegloff, E. A. (2007). *Sequence organization in interaction. A primer in conversation analysis* (Vol. 1). Cambridge University Press.

Skogmyr Marian, K. (2021/this issue). Initiating a complaint: Change over time in French L2 speakers' practices. *Research on Language and Social Interaction*, *54*(2), 163–182. https://doi.org/10.1080/08351813.2021.1899709

Skogmyr Marian, K., & Balaman, U. (2018). L2 interactional competence and its development: An overview of conversation analytic research on interactional change over time. *Language and Linguistics Compass*, *12*(8), 1–12. https://doi.org/10.1111/lnc3.12285

Stevanovic, M. (2018). Social deontics: A nano-level approach to human power play. *Journal for the Theory of Social Behavior*, *48*(3), 369–389. https://doi.org/10.1111/jtsb.12175

Stivers, T. (2004). "No no no" and other types of multiple sayings in social interaction. *Human Communication Research*, *30*(2), 260–293. https://doi.org/10.1111/j.1468-2958.2004.tb00733.x

Thompson, S. A., Fox, B. A., & Couper-Kuhlen, E. (2015). *Grammar in everyday talk: Building responsive actions*. Cambridge University Press.

Tomasello, M. (2003). *Constructing a language: A usage-based approach to child language acquisition*. Harvard University Press.

Wagner, J., Pekarek Doehler, S., & González-Martínez, E. (2018). Longitudinal research on the organization of social interaction: Current developments and methodological challenges. In S. Pekarek Doehler, J. Wagner&, &

E. González-Martínez (Eds.), *Longitudinal studies on the organization of social interaction* (pp. 3–35). Palgrave Macmillan.

West, C. (1990). Not just doctors' orders: Directive-response sequences in patients' visits to women and men physicians. *Discourse & Society*, *1*(1), 85–112. https://doi.org/10.1177/0957926590001001005

Whalen, J., & Zimmerman, D. H. (1998). Observations on the display and management of emotion in naturally occurring activities: The case of" hysteria" in calls to 9-1-1. *Social Psychology Quarterly*, *61*(2), 141–159. https://doi.org/10.2307/2787066

Wootton, A. J. (1997). *Interaction and the development of mind*. Cambridge University Press.