

Article

Multimodal Resolution of Overlapping Talk in Video-Mediated L2 Instruction

Taiane Malabarba ^{1,*}, Anna C. Oliveira Mendes ¹ and Joseane de Souza ² 

¹ Department of English and American Studies, University of Potsdam, 14469 Potsdam, Germany; oliveiramend@uni-potsdam.de

² Pastor Dohms School, Porto Alegre 90550-031, Brazil; joseanesouza@dohms.org.br

* Correspondence: tmalabar@uni-potsdam.de

Abstract: This paper investigates a pervasive phenomenon in video-mediated interaction (VMI), namely, *simultaneous start-ups*, which happen when two speakers produce a turn beginning in overlap. Based on the theoretical and methodological tenets of conversation analysis and interactional linguistics, the present study offers a multimodal and sequential account of how simultaneous start-ups are oriented to and solved in the context of English as an additional language (L2) tutoring. The micro- and sequential analysis of ten hours of screen-recorded video-mediated data from tutoring sessions between an experienced tutor and an advanced-level tutee reveals that the typical overlap resolution trajectory results in the tutor withdrawing from the interactional floor. The same analysis uncovered a range of resources, such as lip pressing and the verbal utterance ‘go ahead’, employed in what we call *enhanced explicitness*, through which the withdrawal is done. The orchestration of these resources allows the tutor to exploit the specific features of the medium to resolve simultaneous start-ups while also supporting the continuation of student talk. We maintain that this practice is used in the service of securing the learner’s interactional space, and consequently in fostering the use of the language being learned. The results of the study help advance current understandings of L2 instructors’ specialized work of managing participation and creating learning opportunities. Being one of the first studies to detail the practices involved in overlap resolution in the micro-context of simultaneous talk on Zoom-based L2 instruction, this study also makes a significant contribution to research on video-mediated instruction and video-mediated interaction more generally.

Keywords: video-mediated interaction (VMI); English as an additional language (L2); teaching; turn-taking; overlap resolution; ‘go ahead’; multimodality; conversation analysis; interactional linguistics



Citation: Malabarba, Taiane, Anna C. Oliveira Mendes, and Joseane de Souza. 2022. Multimodal Resolution of Overlapping Talk in Video-Mediated L2 Instruction. *Languages* 7: 154. <https://doi.org/10.3390/languages7020154>

Academic Editors: Pia Sundqvist, Erica Sandlund, Marie Källkvist, Henrik Gyllstad, Juana M. Licerias and Raquel Fernández Fuertes

Received: 1 November 2021

Accepted: 27 May 2022

Published: 20 June 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

One of the basic rules of the turn-taking system described by Sacks et al. (1974) is that one party talks at a time, with no or minimal gap or overlap between turns. When overlap occurs, participants draw on the *overlap resolution device*, which is an organized set of practices by which the parties attempt to resolve the overlap and return to a ‘one-at-a-time’ situation (Schegloff 2000). Although several kinds of overlapping talk do not need to be avoided or resolved (see, e.g., the case of early responses, explored by Deppermann and Schmidt 2021), certain forms of overlapping talk are considered “recognizable events” (Schegloff 2000, p. 11), which participants orient to and act upon to secure the progressivity of ongoing talk. An example of this type of overlap are *simultaneous start-ups* (Schegloff 2000), which are the focus of this paper. Simultaneous start-ups are overlaps that occur at transition relevance places (TRP), i.e., places in the interaction “where a change of speakership becomes a salient possibility” (Clayman 2012, p. 151) and are recognized as initiating a new turn-at-talk.

In order to avoid beginning a new turn in overlap or to resolve simultaneous start-ups as quickly as possible, speakers monitor each other closely; that is, ending and incoming

turns are projected via participants' gestures, modified posture, facial expressions, and manipulation of objects (Mortensen 2009; Oloff 2013; Mondada 2014, 2007). In video-mediated interaction, however, participants' access to each other's vocal and embodied behavior is impacted due to a range of medium-specific features, such as corporal segmentation, lack of reliance on eye gaze direction, and synchronicity issues, e.g., delays (Luff et al. 2003; Arminen et al. 2016; Licoppe and Morel 2012; Ruhleder and Jordan 2001; Rusk and Pörn 2019; Seuren et al. 2021). As a result, overlapping talk that requires negotiation from the participants (e.g., simultaneous start-ups) has been shown to happen more frequently in video-mediated encounters in comparison to face-to-face interaction (Schneider 2017; Olbertz-Siitonen 2015; Seuren et al. 2021), and is impacted by, e.g., the higher transition time (487 ms) in comparison with face-to-face conversation (135 ms) (see Boland et al.'s (2021) experimental study with dyads interacting via Zoom).

In the data analyzed in this paper, which stem from English tutoring sessions held over Zoom, we identified 41 instances of simultaneous start-ups in ten sixty-minute sessions. Our analysis shows that most cases of this type of overlap result in the student remaining on the floor, and that the tutor draws on a range of multi-semiotic resources to enhance the explicitness of the turn-yielding moves that she uses to secure the floor for the student. We argue that this recurrent resolution trajectory reflects participants' shaped conduct in relation to the affordances of the medium, as well as the tutor's orientation to moments of simultaneous start-ups as emergent language use/learning opportunities (Sert 2017) and her role as an 'interactional manager' (Walsh 2006; Kasper 2004) responsible for crafting such opportunities in situ.

In what follows, we review studies on overlap resolution in video-mediated interaction (VMI), video-mediated additional language (L2) learning, and tutoring, which serve as background to the present contribution.

2. Overlap Resolution in Video-Mediated Interaction

Schegloff (2000) identified distinct phases through which overlaps are solved with reference to the onset of the overlapping talk: the *pre-onset phase*, *post-onset phase*, *post-post-onset phase*, and *post-resolution phase*. During the pre-onset phase, the current speaker may react to their co-participant's incipient attempts to take the floor (for instance, through gaze, pointing, loud in-breaths) and design their utterance to prevent overlap (e.g., by speeding up the pace of talk). In the post-onset phase, speakers adapt to the fact that they are producing talk simultaneously. This adaptation includes halts in the progressivity of talk, usually referred to as "hitches" (e.g., cut-offs) or prosodically marked articulation of talk (e.g., slower tempo), labelled "perturbations" (Schegloff 2000, p. 11). Although hitches and perturbations are found in other interactional environments as well, they acquire specific functions in the context of overlap resolution, as they index that the participants have noticed a problem with progressivity that may impact mutual understanding. In the post-post-onset phase, speakers usually launch what Schegloff calls a "contest" for the floor, by speaking louder, for instance. In the post-resolution phase, the "winner" of the floor adapts back to speaking solo (Schegloff 2000, p. 44). At each beat of overlapping talk, the participants take a stance on what to do next, i.e., to produce the next beat in overlap or to stop talking.

More recently, studies working with video-based data (Heath and Luff 1993; Seuren et al. 2021; Ruhleder and Jordan 2001) have shown how, in VMI, the resolution of overlaps involves dealing with the inherent asymmetries and incongruency in the production and reception of interactional moves. For instance, Seuren et al.'s (2021) study of video-mediated consultations in the UK showed that due to delays between the production of a turn by the speaker and the perception of the same turn by the recipient, participants may perceive moments of actual talk as silence. Accordingly, the practices for the resolution of overlapping talk described above may be delayed for one or both interlocutors, and it may take several turns (and new overlaps) for participants to reach an agreement on who should remain on the floor. In a study of the use of videoconferencing to support remote

teamwork, [Ruhleder and Jordan \(2001\)](#) found that transmission delays impacted the course of interaction but seemed to go unnoticed by the participants, who were not able to identify the source of emerging trouble (but see [Olbertz-Siitonen 2015](#) for partially different results). In such contexts, the timely response of one speaker was perceived as late and in overlap with the continuation of another speaker, who might not have continued if the response had come in on time.

All in all, participants in video-mediated interaction do not seem to acknowledge the existence of latency effects ([Heath and Luff 1993](#); [Seuren et al. 2021](#)); that is, they tend to orient to “two non-mutual realities” as a “shared one” ([Seuren et al. 2021](#), p. 66). At the same time, as [Heath and Luff \(1993\)](#) point out, participants’ conduct *is* affected by such incongruencies, as evidenced, for instance, by the upgrading of certain gestures, especially in cases in which their first attempts at certain interactional moves prove unsuccessful.

Against this background, this study explores turn beginnings that are produced in overlap in the context of L2 video-mediated instruction. We empirically show that in the context of our data, L2 tutoring aimed at advancing learners’ speaking skills, simultaneous start-ups are resolved in a more explicit fashion in comparison to face-to-face interaction as described by [Schegloff](#). The typical resolution trajectory and participants’ moves observed by [Schegloff \(2000\)](#) and others ([Oloff 2013](#); [French and Local 1983](#)), i.e., stopping talking and thus yielding the floor, or continuing speaking (e.g., louder) and thus staying on the floor, are not the norm in our data. Rather, in most of our cases, the negotiation involves dropping out of the floor in a rather upgraded fashion even in moments when the overlap is not persisting. As this practice is pervasively used by the tutor only, we claim that it reflects both participants’ adaptation to the specific features of the medium (as prior VMI studies have identified) and the tutor’s orientation to moments of simultaneous start-ups as locally emerging opportunities for the learner to use the language being learned. A similar claim has been made by [Cancino Avila \(2019\)](#), which, as far as we know, is the only CA study on overlapping talk in language instructional settings. Although his analyses took into consideration different types of overlaps and drew on face-to-face multi-party classroom data, his findings indicated that by providing learners with interactional space in moments of overlap, teachers display classroom interactional competence ([Walsh 2006](#)).

3. Video-Mediated L2 Learning

As pointed out by [González-Lloret \(2015\)](#), micro-analytical studies on video-mediated L2 learning can be categorized into two major strands: descriptive studies (e.g., [Uskokovic and Talehgani-Nikazm 2022](#); [Wigham 2017](#); [Hampel and Stickler 2012](#); [Rusk and Pörn 2019](#); [Dooly and Davitova 2018](#); [Jakonen and Jauni 2021](#); [Balaman 2019](#)) and developmental studies (e.g., [Balaman and Doehler 2022](#); [Doehler and Balaman 2021](#); [Sert 2017](#); [Balaman 2018](#)). Descriptive studies seek to unveil the generic structures of VMI, that is, how participants manage sequential organization, turn-taking, repair, etc., in such settings. For instance, [Uskokovic and Talehgani-Nikazm \(2022\)](#) describe how L2 speakers use a gesture with the index finger alone or in combination with the utterance *ein moment* (‘one moment’ in English) to create space for screen-based word searches in interactions-for-learning between German L1 and L2 speakers. In turn, developmental studies seek to track the development of interactional practices over time. For example, in his study of hinting sequences within a task-based activity, [Balaman \(2018\)](#) shows how a learner increasingly diversifies the resources used to provide hints to her classmates about how to complete the task at hand.

A recurring theme in both strands are participants’ methods for securing and maintaining progressivity. Several studies describe how participants use affordances of the medium to restore progressivity, while others focus more specifically on the emergence of practices to deal with technological constraints. [Dooly and Davitova \(2018\)](#), for example, document how the practices of showing smartphone screens and typing are used by a group of teenage learners to maintain progressivity in the face of communication barriers. [Balaman and Doehler \(2022\)](#) and [Doehler and Balaman \(2021\)](#) describe the development

and routinization of grammatical formats (e.g., *let me* and *let's x*) used to transition into screen-based activities. Finally, in their study of Swedish–Finnish Tandem interactions, [Rusk and Pörn \(2019\)](#) show the local emergence of several strategies to deal with lags, such as an explicit orientation by the L1 speaker towards the L2 speaker's incoming turn.

The current paper is situated within the 'descriptive' strand. It adds to this body of work on video-mediated L2 learning by providing a detailed account of participants' interactional conduct when dealing with turn-taking-related disruptions in progressivity in the understudied context of L2 online tutoring.

Tutoring as an Instructional Activity

EM/CA-inspired studies on tutoring, both in face-to-face and mediated by video, have analyzed the micro-analytic and sequential aspects of this type of instructional activity in various disciplines (see, for example, [DiFelice Box \(2015\)](#) and [Creider \(2020\)](#) on Math tutoring involving children and [Bowden and Svahn's \(2020\)](#) single-case analysis of video-mediated Math homework support with an upper-secondary student). In the context of L2 instruction, the rather small body of micro-analytic studies on tutoring has explored, for example, the use of hand gestures to achieve intersubjectivity ([Belhiah 2013](#)), the sequential organization of openings and closings ([Belhiah 2009](#)), and how advice is negotiated and resisted ([Leyland 2018](#); [Park 2017](#); [Waring 2005](#)). And specifically on online L2 tutoring, an instructional set-up that has grown exponentially as a result of technological advances ([Hamid et al. 2018](#)) maximized by the limitations on co-present encounters imposed by the spread of the COVID-19 pandemic, [Nguyen et al. \(2022\)](#) investigated online search sequences in dyadic Skype interactions between an adult tutee and his L2 tutor. Their paper shows how the word 'corkscrew' becomes a teachable and learnable ([Eskildsen and Majlesi 2018](#)) due to participants' impossibility of using the medium's screen-sharing feature. As their analysis reveals, this constraint led to extensive epistemic negotiation between the participants, which occasioned the emergence of the word and its subsequent use by the L2 learner.

One aspect of L2 tutoring that remains largely overlooked refers to specific practices for turn-taking management, which seem to be particularly relevant in this context due to its less constrained interactional arrangement (as compared to regular classrooms). For example, [Belhiah \(2009\)](#) showed that students and tutors tended to carry out the tutoring business rather collaboratively, and that students' self-initiated turns were frequently welcomed by the tutors. In one of his excerpts, a student's initiating turn overlaps with the teacher's, who, similarly to what we find in our data, promptly yields the floor to the learner. The details of this negotiation for the floor, however, are not addressed in his examination. And indeed, no systematic analysis of overlaps in the context of L2 tutoring has been provided so far.

The current paper thus contributes to filling this gap in the CA L2 instruction literature by exploring overlap resolution in video-mediated encounters involving an experienced L2 instructor and an advanced adult learner engaged in what we call *conversational tutoring*. We use this term, which emerged from our work with the data, to refer to one-on-one tutoring sessions led by an L2 teacher that are not part of an institutionalized course program or fixed-term curriculum. Nonetheless, we consider L2 conversational tutoring to be a teaching–learning arrangement. The term 'conversational' refers both to the stated goals of the encounters (to improve students' speaking skills) as well as to their overall sequential structure, which comprises a great deal of interaction that appears 'more conversational' than typically 'instructional' (see context and Methods).

The main goal of this paper is to explore the embodied means through which overlaps are solved and what affordances such resolutions have for this type of L2 instructional activity. A secondary agenda is to consider what interactional demands engenders the typical resolution design found in our data, i.e., the tutor's use of enhanced explicitness in turn-yielding.

4. Context and Methods

The data analyzed in this study come from a budding corpus of video-mediated language instruction. The interactions took place through Zoom, a video conferencing software program that became popular during the outbreak of the COVID-19 pandemic. The participants—Mari (self-identified as female), a certified and experienced instructor who had taught L2 English for more than twenty years both in language courses and at the university level, and Ivo (self-identified as male), an adult B2-level learner—started meeting weekly in April 2020 (Figure 1). Both participants are Brazilian and were living in Europe at the time of the data collection. Ivo’s stated goal with the tutoring sessions was to improve his English speaking skills.



Figure 1. To the left, Ivo (the tutee); to the right, Mari (the tutor).

As stated earlier, the lessons were not based on any pre-assigned course material and were designed by the tutor as mostly comprising conversational practice that was very seldom interrupted for corrections (most corrections were done after the session via WhatsApp or a shared editable file). When there were specific grammatical or lexical units pre-prepared by the tutor, they often drew on issues that emerged during the sessions or related tasks.

Broadly speaking, Ivo and Mari’s encounters followed an overall sequential structure comprising several phases. These included both the general phases identified in everyday and institutional video-mediated interactions, such as pre-opening/opening and pre-closing/closing phases (Mondada 2015; Ilomäki and Ruusuvuori 2020) as well as typical language instruction phases (e.g., homework checking and task instruction giving), although these more typical ‘instructional’ phases were not present in some of the encounters and were held primarily on a conversational basis. The pre-opening and opening phases comprised checks from both parties regarding the quality of the audio and the video and in which the participants greeted each other. Immediately after this phase, Ivo and Mari engaged in extended *howareyou* sequences and updates, followed by a transition to the ‘business of the day’, proposed by Mari, which included, e.g., a pre-reading activity. During the activities proposed by the instructor, the participants often initiated sequences of talk that were arguably ‘conversational’, i.e., which did not specifically advance the proposed activity.

The recordings were made from Mari’s computer with the written consent of both participants. For the purposes of the study presented here, the first ten recorded lessons were transcribed according to the GAT 2 conventions for English (Couper-Kuhlen and Barth-Weingarten 2011) and the Mondada system for multimodal transcription (Mondada 2019). Combined, these systems allow for detailed prosodic–phonetic and bodily–visual description¹.

This study draws on the methodological framework of Conversation Analysis (CA) and Interactional Linguistics (IL). IL investigates linguistic resources as they are used in social interaction (Couper-Kuhlen and Selting 2018) and has been heavily influenced by CA, which seeks to unveil the nature of the orderliness observed in social interaction. Both fields depart from the basic assumption of “order at all points” (Sacks 1992, p. xlvii), i.e., no detail is dismissed as irrelevant to participants before careful consideration. Likewise, IL and CA approach social interaction from an emic perspective, through which context is reflexively

constructed through participants' talk (Schegloff 1997) and participants' sensemaking practices to accomplish actions are reconstructed through the next-turn proof procedure.

As a CA/IL study, our analysis started with the collection and detailed transcription of candidate single instances of the focal phenomenon (see Hoey and Kendrick 2017 for collections in CA), which yielded a total of 41 clear cases of simultaneous start-ups. We considered turn beginnings that overlapped in the pre-onset and onset phase (see Section 2 above) to be clear cases of simultaneous start-ups. We excluded non-competitive overlaps, such as utterance completions and co-constructions, agreements, brief assessments, and laughter. As we only had the recordings from Mari's perspective, our analysis was based on the 'version' of the interaction uniquely available to one of the participants (cf. Olbertz-Siitonen 2015).

5. Results: The Multimodal Design of Simultaneous Start-Up Resolutions

The analytical part of this paper illustrates the typical simultaneous start-up resolution trajectory found in our data, which comprises Ivo "winning" the floor (Schegloff 2000) and Mari indicating her dropping out through what we call *enhanced explicitness*, i.e., through the combination of different linguistic and embodied resources used in what Mondada (2014) calls a *complex Gestalt*. One case in which the *enhanced explicitness* strategy is not used is presented as well, with the aim of discussing possible explanations for the distribution of the *enhanced explicitness* strategy in our collection.

5.1. The Lip Pressing Gesture

Excerpt 1 shows Mari's use of a lip gesture to more explicitly withdraw from the competition for the floor. It takes place at the beginning of the session. Mari and Ivo had spent a couple of minutes looking up landmarks in Mari's neighborhood, at that time unknown to Ivo, on Google Maps. Mari had told Ivo that one of the advantages of living in that particular area was the easy access to parking slots in comparison to other parts of the city. We join their conversation as Ivo affiliates with this stance. The simultaneous start-up takes place in lines 6 and 7.

Excerpt 1 New buildings - zoom 07.05.20_00:03:39-03:50

```

01  IVO:  °h and uh in ~the oldest center it's uh:
    ivo      ~gazes up/left-->

           ~almost &*imPOSSible;~
    ivo      -->~gazes down-----~
    mar      &contracts closed mouth-->
    mar      *nods-->

02  IVO:  ~i think jus~t the_uhm: &*people that live ~uh °h~ there
    ivo      ~gazes left~gazes at screen-----~left~fwd-->
    mar      -->&*

           ca~n can park on the S~TREETs--~
    ivo      -->~gazes left-----~fwd-----

03  MAR:  ~M hm-
    ivo      ~gazes left-->

04  IVO:  ~ (0.5)
    ivo      -->~gazes right/down-->

05  MAR:  that's ↑~TRU%E==
    ivo      -->~gazes down-->
    ivo      %opens mouth

→ 06      ==#°h ▲ [(a~n)#]

→ 07  IVO:  [ a~nd#] uh: (.) <<h>&#in your REgi▲on*,>
    ivo      -->~gazes fwd~gazes left-->
    mar      ▲moves head to left, gazes twd window-▲head to screen>>
    mar      &presses lips-->
    mar      *nods-->
    fig      #Fig.2      #Fig.3      #Fig.4

```



Figure 2



Figure 3



Figure 4

```

08  IVO:   are &the:re just* ~uh:m: new BUIL~Dings:,
      ivo           -->~gazes fwd-----~gazes down-->
      mar           -->&
      mar           -->*

09           (1.0)

10  MAR:   *YES;
      mar     *nods-->

11           (0.5)

12  MAR:   it's [it's d*~] DIFferent;=

13  IVO:           [YES, *~]
      ivo           -->~gazes fwd-->>
      mar           -->*

14  MAR:   =uh_HUH,=

15           =the buildings are m: MODern;

```

After Ivo's turn (lines 1–2), Mari produces a minimal token that is hearably an agreement with what Ivo just said (line 3). As Ivo gazes down and does not produce another turn (line 4), Mari explicitly formulates her agreement with “that's true” (line 5), which constitutes both a TRP and a place of possible sequence closure. This is when the simultaneous start-up takes place, with both Mari and Ivo producing *and*-prefaced new turns (lines 6 and 7). For her part, when Mari produces the beginning of her turn, she moves her head to the right towards the window. Although we cannot know for sure, this body movement may have contributed to the simultaneous start up, as it may have impeded the monitoring of each other's embodied turn incipency. Right after the overlap, we hear a halt in Mari's turn while she is looking away. For his part, Ivo produces a hesitation marker, *uh:*, followed by a pause, and then continues speaking. Mari's next embodied action, done through a pressing lip gesture (line 7, Figure 4) beginning at the end of Ivo's hitch, confirms her withdrawal from the floor. As Ivo continues speaking, Mari keeps her lips pressed, turns to the screen, and nods.

Because of latency issues, participants in video-mediated interaction may face difficulties signaling to their co-participants that they have dropped out of the floor, which can lead to extended stretches of simultaneous talk and several attempts at overlap resolution (Seuren et al. 2021). In the example shown here, the lip pressing gesture—a more explicit resource compared to glottal or labial stops (Schegloff 2000)—allows Mari to underscore her non-speaking stance, and functions as another signal for Ivo to continue his turn. Excerpt 1 thus shows Mari's exploitation of the talking heads configuration (Licoppe and Morel 2012) afforded by the video-conferencing program. Furthermore, because it is non-vocal, this resource facilitates the coordination of speaker transition (cf. Seuren et al. 2021).

5.2. The Go Ahead Utterance

Excerpt 2 was extracted from Mari and Ivo's first lesson, in which the participants are engaged in a video activity. The video is about strategies that polyglots use to learn a new language. Mari's 'instructional project' (Kimura et al. 2018) seems to be to assess Ivo's understanding of this type of material as well as to discuss language learning strategies that may work for him in particular. The base sequence starts before the excerpt, as Mari pauses the video clip and asks Ivo to report on main ideas conveyed in the video and on whether he agrees with them. We join the conversation towards the end of Ivo's response.

Excerpt 2 Barrier – zoom 15.04.20_01:05:16-01:06:09

```

01  IVO:  uh:m (---) uh: uh no~waDAYS==
      ivo  >>gazes down/left---gazes right-->

02      =wh:: what that i feel with ENGLISH;

03  IVO:  i:s that i: (-) i s~top$ in a:: ((click)) °h with a barRier;
      ivo  -->~gazes fwd-->
      ivo  $RH showing level/barrier-->

04  IVO:  i d~on$ i i don't not~ice that i'm imPROVing,
      ivo  -->~gazes left-----gazes fwd-->
      ivo  -->$

05      * (--)
      mar  *nods-->

06  IVO:  u:h~ for more* than i trie(d) to sp~eak~ with *PERson;=
      ivo  -->~gazes right/down-----fwd-right/down-->
      mar  -->* *nods-->

07  IVO:  =i:: READ-

08  IVO:  i watch M~Ovies- ~
      ivo  -->~gazes fwd~

09      ~u[h * i ~] feel* that i (stop) in THIS-

10  MAR:  [m_*HM, ~]
      ivo  ~gazes right~gazes fwd-->
      mar  *nods slightly*

11  IVO:  and si:: can't improve in more tha ~more than *THIS;
      ivo  -->~gazes right/down-->1.18
      mar  $gazes down, takes notes-->1.20
      mar  -->*

12  MAR:  %m_ΔHM,
      ivo  %opens mouth-->
      ivo  Δbreathes in moving torso slightly upwards-->

13      (1.0)% (0.7)
      ivo  -->%closes mouth and opens it again-->

14  IVO:  <<creaky>Δu:[::h> %]
      ivo  -->Δ

15  MAR:  [<<p, len>m_HM,>%]
      ivo  -->%

16      (-)

17  MAR:  ((click)) °h [<<all>#yEs;ε>]
      mar  %keeps mouth slightly open-->

18  IVO:  [ #so: ] <<creaky>uh~ uh>%
      ivo  -->~gazes fwd-->
           %closes mouth-->

```



Figure 5



Figure 6

```

19      * (--)#
mar     *nods-->
fig     #Fig.6

20  MAR:    &Bgo aHEAD;*
mar     -->&
mar     -->Bgazes fwd-->1.29
mar     -->*

21      <<creaky>%SORry.>&
mar     &smiles-->
ivo     -->%

22  IVO:    yeah~ so so i i i &i don't know if i have~ uh:m (a ob)
ivo     -->~gazes left-----~gazes fwd-->
mar     -->&

        uh:~m (.)
ivo     -->~gazes diffusely left/down/right-->

23      something that i can: (-) transpass this barRIer.=

24      =because uh: now that i (.) that i feel that i (stopping/ stop
        in) in this LLevel;=

25      =(an`/when) (2.2) i can't imPROVE <<laughing>(more than
        [thi]s.)>

26  MAR:    [(a)]

27      ~uh_HUH,=
ivo     -->~gazes fwd-->>

28      =it's hard for you Bto to go FURTHER;
mar     -->Bgazes down-->>

29      RIGHT,

```

This excerpt exhibits two simultaneous start-ups (lines 14–15 and 17–18). While our focus is on the second one (line 17), we will begin by taking a close look at the first due to its import to the emergence of the second simultaneous start-up. Between lines 1 and 11, Ivo describes his current difficulties with learning English. Although Ivo's turn "and i can't improve in more tha more than this" (line 11) is syntactic and pragmatically complete, its mid-rising final intonation while keeping his gaze down suggests more to come. Throughout Ivo's TCU in line 11, Mari is looking down, involved with note-taking. Mari utters a mid-rising continuer, *m_HM*, (line 12), thereby registering Ivo's TCU and

treating it as a non-turn-final one (Jefferson 1984). It is during a pause of 1.7 s (line 13), when Mari is taking notes and Ivo is looking down, that the first simultaneous start-up takes place (lines 14–15).

During this pause, had Mari been looking at her screen she might have noticed Ivo's visible in-breath (through torso movement) and mouth opening, both signs that he was about to start talking again. Ivo then utters a pre-turn-beginning *u:h* (line 14) (Schegloff 2010), potentially not noticed by Mari, which indicates an attempt to secure his turn space while he formulates his utterance. Mari comes in in overlap with a second continuer "m_HM", with reduced loudness and slower tempo (line 15), which does the work of explicitly displaying attentiveness to Ivo's ongoing turn while she is gazing down engaged in note taking. As Ivo does not launch a TCU right away, as projected by the *u:h* (note the pause in line 16), another simultaneous start-up ensues. While Ivo finally launches his TCU with *so* (line 18), Mari produces a click and delivers a *yes* with mid-falling final intonation (line 17), after which Mari's mouth is kept slightly open, indicating that Mari was going to continue speaking. This conduct seemingly indicates both that she was done note taking and her readiness to take over speakership (Jefferson 1984).

At this point, Ivo produces two *uhs* (line 18) and stops talking, which hearably constitute overlap-related hitches (Schegloff 2000). That Mari interprets Ivo's conduct as somewhat oriented to the overlap as "a recognizable event" (Schegloff 2000, p. 11) is supported by Mari's subsequent conduct. For one, she closes her mouth and nods (lines 18–19). Second, as Ivo does not continue immediately (see the pause in line 19), Mari utters a *go ahead* while shifting her body positioning and gaze direction (line 20). The verbal *go ahead* is followed by an apology (line 21), and a smile (line 21). In her move to make sure that Ivo remains on the floor, similarly to what she does in Excerpt 1, Mari uses the enhanced-explicitness strategy in the resolution of the simultaneous start-up. This time, such enhanced explicitness is embodied by the use of a verbal utterance, through which Mari officially allocates and secures the floor to Ivo and encourages him to continue.

Locally, two main contingencies may explain the design of her withdrawing. First, this is a second simultaneous start-up that happens right after the first one. Second, Ivo stops talking after uttering two non-lexical *uh*. Arguably, those two events together render this moment as one of persisting turn-taking issues, which warrants the use of even more explicit means (as compared to Excerpt 1) to resolve the overlap. As Mari turns to the screen while saying "go ahead", she can see that Ivo adopts a stand-by posture as he stops talking and stares at the screen (line 18, Figure 6). The apology "sorry" (line 21) further underscores the inappositeness of Mari's incoming and indexes that Mari holds herself accountable, as the one responsible for the smooth unfolding of the interaction, for not paying full attention to Ivo's moves. This orientation to Mari's role as interactional manager and instructor is further showcased by the Excerpt 3 below.

5.3. The Combination of the Lip Pressing Gesture and Go Ahead Utterance

Our third example comes from Mari and Ivo's first session, in April 2020, when local COVID-19 lockdown measures were slowly starting to be lifted after a stricter lockdown period. Mari and Ivo had been talking about how they felt during the lockdown. The simultaneous start-up is located in lines 7 and 8.

Excerpt 3 Fresh air – zoom 15.04_00:04:12-00:04:35

```

01  MAR:  bu:t uh yes i miss the fresh AIR (-) •as well.
      mar                                     •raises EB-->1.09

02      (-)

03  MAR:  i really βDO;=
      mar                                     βgazes down-->

04  IVO:  =m_HM,

05  MAR:  ((click))but=(.) that's IT==rIght;β
      mar                                     -->βgazes to screen-->

06      that's what we HAVE for nOw.=

→ 07      =°h βa:nd [eh:mβ]

→ 08  IVO:      [°h β] ~&<<h>'▲yEah~> b%+
      ivo      ~blinks RE--~
      ivo      %closes mouth-->
      ivo      +nods slightly-->
      mar      -->βgazes downβ
      mar      &opens mouth-->
      mar      ▲head backward-->

09      (0.3)▲(0.2) &•# (0.3) ▲ & +
      mar      -->•
      mar      -->&presses lips&decompresses lips-->
      mar      -->▲.....▲head back to center-->>
      ivo      -->+
      fig      #Fig.7

```



Figure 7

```

10  MAR:    *•&go aH*E•ADh.&
      mar    *nods-----*
      mar    *raises EB*
      mar    -->&                &smiles-->

11  IVO:    %+<<all>oKAY,>&+
      ivo    -->%
      ivo    +nods-----+
      mar    -->&keeps smiling, squints eyes ('eye smile')-->

12      (.)

13      ~+<<laughing>n oh #o>KAY;~+
      ivo    ~rolls eyes-----~
      ivo    +throws head back-----+
      fig    #Fig.8

```



Figure 8

```

14  IVO:    but eh i i &(would to) suggest SOMETHing;&
      mar    --->&/////////////////////////&

15      because eh near to my place there is

      tEjo is eh *very CLOSE;
      mar    *nods----->>

```

Lines 1–3 show Mari’s resaying of her previous comment about missing the possibility of spending time outdoors. The design of Mari’s turn in line 6, along with the alternating eye gaze direction (lines 3–7), suggests a potential sequence and topic-closure (Schegloff 2007). In line 7, Mari’s turn is delivered as she is gazing down (potentially at her notes), which, together with the connector *and*, suggests that she was indeed about to launch a new (sub-)topic. As Mari is gazing down, Ivo opens his mouth and then produces an audible in-breath, which is delivered in overlap with Mari’s turn.

Mari’s dropping out of the floor happens as soon as she hears Ivo’s turn. She first does this by simply interrupting her talk. As a consequence, the continuation of Ivo’s turn, containing a prosodically marked high-pitched *yeah* and presumably the first sound of the conjunction *but*, is heard in the clear (line 8). This turn design projects a concessive *yes-but* construction (Couper-Kuhlen and Thompson 2000), and thus suggests that Ivo’s talk was going to propose a slightly different understanding than the one formulated by Mari in lines 5–6, while remaining on the same topic.

The halt of his turn seemingly displays Ivo’s stance towards the fact that he and Mari are talking in overlap. He does not compete for the floor and seems to be uncertain about who should continue, which causes the talk to be momentarily halted. During this pause (line 9), Mari’s mouth is closed and Ivo is nodding slightly. Mari then further displays her withdrawal by pressing her lips while moving her head back to the center/front of the screen, which underscores her readiness to listen to Ivo. What follows is a yet more explicit display of Mari’s withdrawal: she produces a turn containing *go ahead* while nodding and raising her eyebrows, then smiles (line 10). For his part, Ivo first produces a free-standing *okay* (Couper-Kuhlen 2021), accompanied by a subtle nod (line 11). Although *okays* are particularly known across languages for acknowledging receipt of information (Couper-Kuhlen 2021; Oloff 2019), the prosodic delivery (faster than surrounding talk and

with mid-rising intonation) of Ivo's *okay* suggests that it may be doing more than merely acknowledging that Ivo has heard that Mari and he are speaking in overlap. However, in the specific context of simultaneous start-up resolution in video-mediated interaction, it is difficult to determine its function. It could be checking whether the lag was over on Mari's end, requesting confirmation from Mari that he could/should continue, or even yielding the floor to Mari. Finally, through its strategic position and design, Mari's smile does several things. First, it further signals her withdrawal from the floor, and thus from the negotiation of speakership. Second, by maintaining her smile at this point in the interaction, Mari imprints a playful stance towards her invoking of her primary rights over the interactional floor as well as the tacit norms of their encounter, which includes Mari facilitating Ivo's access to the floor.

The design of Mari's actions via a cluster of facial gestures (Kendon 2004) and other linguistic and prosodic resources in the overlap resolution in this case points to the local management of the business at hand, i.e., managing the interactional floor and maximizing learners' language use by contingently securing the floor to the student in moments of overlap. By not continuing to speak and signaling her dropping out of the floor in such a marked fashion, Mari momentarily invokes her deontic status, i.e., "the relative position of power that a participant is considered to have or not to have, irrespective of what he or she publicly claims" (Stevanovic 2018, p. 375), as instructor. Ivo's subsequent turn with another *okay*, this time prefaced by a change of state token (Heritage 2016) *oh* (line 13), is delivered with mid-falling intonation in a multimodal package of its own, with laugh particles and accompanied by an eye roll and a head movement (line 13, Figure 8), thus affiliating with the playful stance projected by Mari's embodied actions to resolve the overlap. It complies with Mari's turn-yielding move in a humorous enactment of feigned resistance.

Excerpts 1–3 showcase, in micro-analytical and sequential terms, a recurrent trajectory of simultaneous start-up resolution in our data, i.e., that the tutor orients to the tutee as the one who should continue talking after moments of simultaneous start-ups. They are also representative of how the handing over of the floor happens; it is accomplished through a rich multimodal *Gestalt* of resources mobilized by the tutor in a locally-sensitive way as she adjusts her conduct to the medium as well as to her student's moves. We propose that such moves, i.e., recurrently dropping out of the floor at moments of overlap and indicating it through more explicit embodied and linguistic methods, constitute an emergent interactional practice used by the teacher to manage turn-taking and yield the turn to the student. As such, we argue that this recurrent resolution trajectory reflects and constitutes part of the tutor's role of 'interactional manager' (Walsh 2006; Kasper 2004) and as the crafter of learners' emergent language use opportunities (Sert 2017).

That this practice reflects the tutor's orientation to her instructional role and the larger agenda of securing and maximizing the student's interactional space is further substantiated by an example in which the more explicit resources are not employed. We turn to this last example in Excerpt 4 below.

5.4. A Contrasting Case: The Resolution of Overlap without the Enhanced Explicitness Strategy

Excerpt 4 is a case in which the simultaneous start-up is solved *without* the enhanced explicitness strategy. It takes place half-way through the session. As Mari corrected Ivo's use of pronouns, they ended up talking about the adequacy of pronouns and their relation to sexist language. We join the interaction when Mari is telling Ivo about her M.A. dissertation, which she wrote in Portuguese. In line 1, Mari states that when writing her dissertation, she decided to use both the masculine and the feminine nouns to refer to groups that include both men and women, although mixed groups are commonly referred to by the use of masculine forms in Portuguese. This is what the indexical *that* in line 1 refers to.

Excerpt 4 Dissertation – zoom 22.04.20_00:36:19-00:36:36

01 **MAR:** \$but (.)\$for my dissertation i deci\$ded# to do THAT;=
 mar \$.....\$RH down, fingers together-\$closes RH, keeps at
 shoulder height-->1.08
 fig #Fig.9



Figure 9

02 =because i was writing about GENDER (.) as well;

03 (-)

04 **MAR:** |so i thought it w[as imPORT]a[nt;]
 scr |ivo: image freezes-->

05 **IVO:** [m_HM,]

06 [AH-]

07 *(0.5)|~ (0.6)*
 scr -->|ivo: image is back to normal-->1.16
 ivo ~raises eyebrows-->
 mar *nods-----*

→ 08 **MAR:** \$an[d #lang]

→ 09 **IVO:** [it's #a~bout] *wh\$At*\$ your disserTA~#tion,
 ivo -->~
 ivo \$puts hands on earphones--->
 ivo ~looks down/left->
 mar -->\$moves RH fwd, palm up\$retracts, rests hand on shoulder-->
 mar *nods-
 fig #Fig.10 #Fig.11



Figure 10



Figure 11

```

10  MAR:  (.)
11  IVO:  (hm) $~ [(ts)]
      ivo  -->$
      ivo  -->~gazes to screen-->>
12  MAR:  [ A]H:;
13  Bmy disserBtation was aBbou:t Buh
      mar  Bgazes leftBgazes fwd---Bgazes upBgazes fwd-->>
14  have you ever heard of ((says name of Brazilian institution));
15  (1.4)
16  IVO:  |+YES $yes;
      scr  -->|ivo: image is delayed in relation to sound-->>
      ivo  +nods-->>
      mar  -->$moves hand to the side of head-->>

```

Mari's telling about her dissertation is received with a continuer (line 5) followed by a change-of-state token (line 6). This change-of-state token (Heritage 2016), along with Ivo's raised eyebrows, displays Ivo's surprise on learning about Mari's dissertation topic, one Ivo, as he himself had stated previously, was interested in. During a pause of one and a half seconds (line 7), Mari seems to be inviting elaboration related to Ivo's display of surprise, which is indicated by her silence and nods. At the same time, Mari could be waiting for a transmission problem to be resolved, as Ivo's turn (lines 5–6) can only be heard, not seen (see Ivo's image freezing in line 4). After this silence, Ivo makes his interest on the topic of Mari's dissertation more evident through the initiation of a new sequence projected by a question (line 9). This new sequence clashes with a potential topic expansion by Mari (line 8), producing a simultaneous start up. The content of Mari's turn beginning suggests that she was going to advance the larger topic of the interaction at that point (sexism and language), instead of producing further talk on her dissertation.

As with most of the cases in our collection, what follows the overlapping talk is Mari's ongoing turn suspension. However, the design of her dropping out this time is considerably different from the previous examples. In Excerpt 4, the simultaneous start-up is resolved through the deployment of the most basic resource available to speakers in talk-in-interaction, i.e., to simply stop talking. The action of dropping out of the floor and thus yielding the floor to Ivo is not highlighted by the lip gesture nor does Mari employ the utterance *go ahead*. It is comprised solely of stopping the production of her turn, nodding, and halting her right-hand movement and its retraction to her shoulder, where she lets her hand rest (line 9). This movement signals her withdrawal of the floor and indexes her reciprocity state (Oloff 2013).

While the design of Mari's withdrawal in this occurrence (halt of talk plus the hand gesture retraction) allows Ivo to continue (as the other designs do), it does this in a way that does not impose on him in the same manner as a fiercely closed mouth with pressing lips or a verbal *go ahead* would. Seemingly, in this particular example, as well as in the other two examples found so far in our data in which Mari is the one talking about personal affairs, Mari's role of interactional manager/instructor is backgrounded; that is, momentarily, Mari and Ivo are not primarily oriented to their deontically asymmetric roles, but rather to their roles as unacquainted interactional partners (Svennevig 2000) who are currently

engaged in learning more about each other's personal affairs. This example thus supports the claim that the backgrounding and foregrounding of the tutor's role of fostering learner participation is constituted by and reflected in how the simultaneous start-ups are oriented to and solved.

6. Discussion and Implications: Overlap Resolution and the Management of Language Use/Learning Opportunities

This paper has examined the resolution trajectory of simultaneous start-ups in video-mediated L2 tutoring. With a decided focus on the cases in which the tutor withdraws and yields the floor to the student, we observed the mobilization of a set of resources comprising the timely termination of talk, mouth and lip gestures, smiles, hand gestures, and the use of the expression *go ahead*. We have shown that these resources, locally used to more explicitly mark the tutor's withdrawal from the floor, support the student's incipient talk and reflect how participants' conduct is shaped by the medium. Furthermore, we have argued that this practice of enhanced explicitness secures the learner's interactional space in moments of overlap in VMI, and thereby indexes the tutor's role as the *craft*er of emergent language-use opportunities.

The present paper has implications for both VMI and for CA-SLA research. Our analysis has unveiled the more prominent function that certain resources acquire in VMI. For example, we have shown how pressing the lips—a practice that has not been reported by previous studies in relation to overlapping talk—is recurrently used in our data. Through the prominence of her lip pressing gesture after dropping out of the floor, the tutor secures her recipient mode and makes use of her rights as interactional manager to influence the course of the overlap resolution. This finding supports earlier research (e.g., [Heath and Luff 1993](#); [Licoppe and Morel 2012](#)) in suggesting that facial gestures gain prominence over hand gestures in VMI, which is an example of how participants exploit the 'talking heads' configuration ([Licoppe and Morel 2012](#)).

As this study is limited to one dyad, more research is needed in order to determine whether the enhanced explicitness strategy is likely to be encountered in interactions involving other conversational partners in VMI. Our analyses suggest that participants' deontic status may play a role in determining who can use the enhanced explicitness strategy for overlap resolution. Further research could verify whether this holds true for other contexts and with participants with more symmetrical deontic statuses. It would also be relevant to verify whether simultaneous start-ups are resolved differently over time, as Mari and Ivo become more familiar with the medium and more acquainted with each other. Future research could also investigate which other practices are used by L2 instructors and participants in VMI more generally to solve simultaneous start-ups as well as other types of interactional hurdles.

Another avenue for future VMI studies is the investigation of how micro-sequentiality ([Deppermann et al. 2021](#)) is accomplished in video-mediated interaction. The notion of 'micro-sequentiality' is a recent development in multimodal CA, and seems crucial to explaining participants' orientations to bodily-visual behaviors ([Mondada 2014](#); [Deppermann et al. 2021](#)). Cases such as the one shown in Excerpt 3 suggest that linguistic and embodied resources may be mobilized across an increasing scale of 'explicitness' (from simply stopping talking to uttering the expression 'go ahead') in order to allow participants to micro-adjust to each other's conduct. This mechanism should be further explored by multimodal conversational-analytical studies on VMI.

Our study contributes new understandings to CA-SLA research. Previous studies have claimed that language use is "the driving force for language development and language learning" ([Eskildsen 2020](#), p. 59) and that an integral part of the specialized work that teachers do involves successfully managing learners' initiatives and creating language learning opportunities (e.g., [Waring 2011](#); [Girgin and Brandt 2020](#); [Sert 2017](#)). Our close analyses of simultaneous start-ups illustrate how such management of learners' self-initiated turns and crafting of language use opportunities is done in the understudied context of adult

L2 tutoring with advanced learners. For one, it invites us to reconceptualize our understanding of ‘language teaching’. In the context of L2 conversational tutoring, juggling the institutional roles of L2 teacher and student and that of conversational partner seems to be a pervasive concern for the participants. As the encounters are to a great extent designed to resemble everyday conversation, a calibration between being a teacher, with all the rights and obligations attached to it, and being a conversational partner, is presumably needed. Indeed, we observe that the categories of teacher and student may be brought to the surface of talk during moments when the progressivity of ongoing talk is at stake, which is the case with simultaneous start-ups. The negotiation that these events require, as the analysis of excerpts 1–3 shows, mobilize a set of practices that seem to actualize the institutional roles of teacher and student in what appear at first glance to be moments of everyday interaction with no fixed instructional agenda (e.g., a grammar point) in place.

All in all, our findings support the general claim that teachers are the “key designers” (Hall 2020, p. 12) of learning opportunities and reveal new facets of the embodied work of teaching (Hall and Looney 2019) that is mobilized in the increasingly popular video-mediated L2 settings. Our findings also provide further evidence for the claim that a core principle of teacher–student interaction is adapting ongoing talk to interactional contingencies (Malabarba 2015; Waring 2016).

In line with the goal of “[m]aking visible the practices and actions of how L2 teaching is accomplished” (Hall et al. 2020, p. 36) in order to produce useful insights to L2 instructors, these findings could integrate evidence-based teacher reflection and professionalization programs (see Hall et al. 2020; Glaser et al. 2019; Sert 2021; Ekin et al. 2021) to leverage L2 instructors’ *e-classroom interactional competence* (Moorhouse et al. 2021). Specifically, our fine-grained analyses of how simultaneous start-ups are solved could help raise L2 instructors’ awareness of the mechanics of dyadic L2 video-mediated instruction. Instructors who are less familiar with VMI should find these results relevant to understanding how pedagogical actions, such as leaving the floor to students in moments of overlap in video-mediated instruction, are accomplished by experienced instructors.

Author Contributions: Data curation, T.M.; manuscript conceptualization, methodology, analysis, original draft preparation, T.M., A.C.O.M., J.d.S.; writing—review and editing, T.M., A.C.O.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed written consent was obtained from the tutor and the tutee involved in this study.

Data Availability Statement: Due to privacy concerns, the data are not available to other researchers.

Acknowledgments: This paper began to gain shape in informal weekly data sessions held over Zoom during the COVID-19 pandemic. A big thanks to Revert Klattenberg, who joined the initial sessions and the analytical brainstorming that culminated in this paper. Our initial analyses were then sharpened with the help of the members of the *Gesprächsanalyse in der Lehrer*innenbildung* (GeLb) group and our colleagues from the Center for Interactional Linguistics at the University of Potsdam in two separate data sessions in which the data and the transcripts were discussed at length. The editors of this Special Issue, Pia Sundqvist, Erica Sandlund, Marie Källkvist, Henry Gyllstad, Juana M. Liceras and Raquel Fernández Fuertes, have also done a brilliant job with helping us reach greater clarity in framing our contribution, sharpening our analysis, and interpreting our findings. And we are indebted to Christl Langer and Marie Will (University of Potsdam) for their invaluable support with the formatting of the paper. Finally, the first author would like to thank the tutor and the tutee for making their sessions available for scientific scrutiny.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Transcription Conventions

Verbal transcription according to GAT 2 (Couper-Kuhlen and Barth-Weingarten 2011)

SYLlable	focus accent
sYllable	secondary accent
?	rising to high
,	rising to mid
-	level
;	falling to mid
.	falling to low
:	lengthening, by about 0.2–0.5 s
::	lengthening, by about 0.5–0.8 s
:::	lengthening, by about 0.8–1.0 s
°h	inbreath of appr. 0.2–0.5 s duration
(.)	micro pause, estimated, up to 0.2 s duration appr.
(-)	short estimated pause of appr. 0.2–0.5 s duration
(-)	intermediary estimated pause of appr. 0.5–0.8 s duration
(--)	longer estimated pause of appr. 0.8–1.0 s duration
(0.5)/(2.0)	measured pause of appr. 0.5/2.0 s duration (to tenth of a second)
[]	overlap and simultaneous talk
[]	
=	fast, immediate continuation with a new turn or segment (latching)
and_uh	cliticizations within units
<<laughing> >	laughter particles accompanying speech with indication of scope
((cough))	non-verbal vocal actions and events
()	unintelligible passage
(may i)	assumed wording
(may i say/let us say)	possible alternatives
<<creaky> >	glottalized
<<h> >	higher pitch register
<<all> >	allegro, fast
'S0	rising accent pitch movement

Multimodal transcription adapted from Mondada (2019)

* *	Descriptions of embodied actions are delimited between two identical symbols that are synchronized with correspondent stretches of talk or time indications.
*-->	The action described continues across subsequent lines
-->*	until the same symbol is reached.
-->1.1	The end of the described action is located several lines after its beginning. It is indicated with the corresponding line number after the first arrow.
>>	The action described begins before the excerpt's beginning.
-->>	The action described continues after the excerpt's end.
.....	Action's preparation.
--	Action's apex is reached and maintained.
,,,,	Action's retraction.
fig	The exact moment at which a screen shot has been taken
#	is indicated with a sign (#) showing its position within the turn/a time measure.

Multimodal transcription symbols:

\$	Mari's hand movements
§	Ivo's hand movements
+	Ivo's nodding
▲	Mari's head direction
*	Mari's nodding
§	Mari's eye gaze
●	Mari's eyebrows
~	Ivo's eyebrow movements and gaze
&	Mari's mouth, lips movements and smile
%	Ivo's mouth movements
@	Mari's chin
Δ	Ivo's torso movements
name :	Indication of video problems due to connection difficulties stating the name of the person whose video is concerned.

Multimodal transcription abbreviations

EB	eyebrows
RE	right eye
RH	right hand
fwd	forward
twd	toward
scr	screen

Note

¹ A complete list of the transcription conventions is found in the Appendix A.

References

- Arminen, Ilkka, Christian Licoppe, and Anna Spagnolli. 2016. Respecifying mediated interaction. *Research on Language and Social Interaction* 49: 290–309. [CrossRef]
- Balaman, Ufuk. 2018. Task-induced development of hinting behaviors in online task-oriented L2 interaction. *Language Learning & Technology* 22: 95–115.
- Balaman, Ufuk. 2019. Sequential organization of hinting in online task-oriented L2 interaction. *Text & Talk* 39: 511–34. [CrossRef]
- Balaman, Ufuk, and Simona Pekarek Doehler. 2022. Navigating the complex social ecology of screen-based activity in video-mediated interaction. *Pragmatics* 32: 54–79. [CrossRef]
- Belhiah, Hassan. 2009. Tutoring as an embodied activity: How speech, gaze and body orientation are coordinated to conduct ESL tutorial business. *Journal of Pragmatics* 41: 829–41. [CrossRef]
- Belhiah, Hassan. 2013. Gesture as a resource for intersubjectivity in second-language learning situations. *Classroom Discourse* 4: 111–29. [CrossRef]
- Boland, Julie E., Pedro Fonseca, Ilana Mermelstein, and Myles Williamson. 2021. Zoom disrupts the rhythm of conversation. *Journal of Experimental Psychology: General*. Advance online publication. [CrossRef]
- Bowden, Helen Melander, and Johanna Svahn. 2020. Collaborative work on an online platform in video-mediated homework support. *Social Interaction. Video-Based Studies of Human Sociality* 3: 3. [CrossRef]
- Cancino Avila, Marco Octavio. 2019. Exploring teachers' and learners' overlapped turns in the language classroom: Implications for classroom interactional competence. *Studies in Second Language Learning and Teaching* 9: 581–606. [CrossRef]
- Clayman, Steven E. 2012. Turn-Constructional Units and the Transition-Relevance Place. In *The Handbook of Conversation Analysis*. Edited by Jack Sidnell and Tanya Stivers. Oxford: Wiley-Blackwell, pp. 150–66. [CrossRef]
- Couper-Kuhlen, Elizabeth. 2021. Language over time: Some old and new uses of OKAY in American English. *Interactional Linguistics* 1: 33–63. [CrossRef]
- Couper-Kuhlen, Elizabeth, and Dagmar Barth-Weingarten. 2011. A system for transcribing talk-in-interaction: GAT 2 for English. *Gesprächsforschung—Online-Zeitschrift zur verbalen Interaktion* 12: 1–51. Available online: <http://www.gespraechsforschung-online.de/fileadmin/dateien/heft2011/px-gat2-englisch.pdf> (accessed on 31 October 2021).
- Couper-Kuhlen, Elizabeth, and Margret Selting. 2018. *Interactional Linguistics: Studying Language in Social Interaction*. Cambridge: Cambridge University Press.
- Couper-Kuhlen, Elizabeth, and Sandra A. Thompson. 2000. Concessive patterns in conversation. In *Cause, Condition, Concession, Contrast: Cognitive and Discourse Perspectives*. Edited by Elizabeth Couper-Kuhlen and Bernd Kortmann. Berlin: Mouton de Gruyter, pp. 381–410. [CrossRef]

- Creider, Sarah Chepkirui. 2020. Student talk as a resource: Integrating conflicting agendas in math tutoring sessions. *Linguistics and Education* 58: 100822. [\[CrossRef\]](#)
- Deppermann, Arnulf, and Axel Schmidt. 2021. Micro-Sequential Coordination in Early Responses. *Discourse Processes* 58: 372–96. [\[CrossRef\]](#)
- Deppermann, Arnulf, Lorenza Mondada, and Simona Pekarek Doehler. 2021. Early Responses: An Introduction. *Discourse Processes* 58: 293–307. [\[CrossRef\]](#)
- DiFelice Box, Catherine. 2015. When and how students take the reins: Specifying learner initiatives in tutoring sessions with preschool-aged children. *Applied Linguistics Review* 6: 23–48. [\[CrossRef\]](#)
- Doehler, Simona Pekarek, and Ufuk Balaman. 2021. The Routinization of Grammar as a Social Action Format: A Longitudinal Study of Video-Mediated Interactions. *Research on Language and Social Interaction* 54: 183–202. [\[CrossRef\]](#)
- Dooly, Melinda, and Nuriya Davitova. 2018. What Can We Do to Talk More: Analysing Language Learners Online Interaction. *Hacettepe University Journal of Education* 33: 215–37. [\[CrossRef\]](#)
- Ekin, Semih, Ufuk Balaman, and Fatma Badem-Korkmaz. 2021. Tracking telecollaborative tasks through design, feedback, implementation, and reflection processes in pre-service language teacher education. *Applied Linguistics Review*, 1–30. [\[CrossRef\]](#)
- Eskildsen, Søren Wind. 2020. From Constructions to Social Action: The Substance of English and Its Learning from an Interactional Usage-Based Perspective. In *Ontologies of English*, 1st ed. Edited by Christopher J. Hall and Rachel Wicaksono. Cambridge: Cambridge University Press, pp. 59–79. [\[CrossRef\]](#)
- Eskildsen, Søren Wind, and Ali Reza Majlesi. 2018. Learnables and Teachables in Second Language Talk: Advancing a Social Reconceptualization of Central SLA Tenets. Introduction to the Special Issue. *The Modern Language Journal* 102: 3–10. [\[CrossRef\]](#)
- French, Peter, and John Local. 1983. Turn-competitive incomings. *Journal of Pragmatics* 7: 17–38. [\[CrossRef\]](#)
- Girgin, Ufuk, and Adam Brandt. 2020. Creating space for learning through ‘Mm hm’ in a L2 classroom: Implications for L2 classroom interactional competence. *Classroom Discourse* 11: 61–79. [\[CrossRef\]](#)
- Glaser, Karen, Maxi Kupetz, and Hie-Jung You. 2019. ‘Embracing social interaction in the L2 classroom: Perspectives for language teacher education’—an introduction. *Classroom Discourse* 10: 1–9. [\[CrossRef\]](#)
- González-Lloret, Marta. 2015. Conversation analysis in Computer-assisted Language Learning. *CALICO Journal* 32: 569–94. [\[CrossRef\]](#)
- Hall, Joan Kelly. 2020. L2 classroom interaction and its links to L2 learners’ developing L2 linguistic repertoires: A research agenda. *Language Teaching*, 1–16. [\[CrossRef\]](#)
- Hall, Joan Kelly, and Stephen Daniel Looney. 2019. Introduction: The Embodied Work of Teaching. In *The Embodied Work of Teaching*. Edited by Joan Kelly Hall and Stephen Daniel Looney. Bristol: Blue Ridge Summit: Multilingual Matters, pp. 1–14. [\[CrossRef\]](#)
- Hall, Joan Kelly, Tianfang Wang, and Su Yin Khor. 2020. The Links between the Linguistic Designs of L2 Teacher Questions and the Student Responses They Engender. *Iranian Journal of Language Teaching Research* 8: 25–40. [\[CrossRef\]](#)
- Hamid, M. Obaidul, Asaduzzaman Khan, and M. Monjurul Islam. 2018. The spread of private tutoring in English in developing societies: Exploring students’ perceptions. *Discourse: Studies in the Cultural Politics of Education* 39: 868–86. [\[CrossRef\]](#)
- Hampel, Regine, and Ursula Stickler. 2012. The use of videoconferencing to support multimodal interaction in an online language classroom. *ReCALL* 24: 116–37. [\[CrossRef\]](#)
- Heath, Christian, and Paul Luff. 1993. Media Space and Communicative Asymmetries: Preliminary Observations of Video-Mediated Interaction. *Human-Computer Interaction* 7: 315–46. [\[CrossRef\]](#)
- Heritage, John. 2016. On the diversity of ‘changes of state’ and their indices. *Journal of Pragmatics* 104: 207–10. [\[CrossRef\]](#)
- Hoey, Elliott M., and Robin Kendrick. 2017. Conversation Analysis. In *Research Methods in Psycholinguistics and the Neurobiology of Language: A Practical Guide*. Edited by Annette M. B. de Groot and Peter Hagoort. Hoboken and Oxford: Wiley-Blackwell, pp. 151–73.
- Ilomäki, Sakari, and Johanna Ruusuvuori. 2020. From appearances to disengagements: Openings and closings in video-mediated tele-homecare encounters. *Social Interaction: Video-Based Studies of Human Sociality* 3: 1–31. [\[CrossRef\]](#)
- Jakonen, Teppo, and Heidi Jauni. 2021. Mediated learning materials: Visibility checks in telepresence robot mediated classroom interaction. *Classroom Discourse* 12: 121–45. [\[CrossRef\]](#)
- Jefferson, Gail. 1984. Notes on a systematic deployment of the acknowledgement tokens “Yeah”; and “Mm Hm”. *Paper in Linguistics* 17: 197–216. [\[CrossRef\]](#)
- Kasper, Gabriele. 2004. Participant Orientations in German Conversation-for-Learning. *The Modern Language Journal* 88: 551–67. [\[CrossRef\]](#)
- Kendon, Adam. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Kimura, Daisuke, Taiane Malabarba, and Joan Kelly Hall. 2018. Data collection considerations for classroom interaction research: A conversation analytic perspective. *Classroom Discourse* 9: 185–204. [\[CrossRef\]](#)
- Leyland, Christopher. 2018. Resistance as a resource for achieving consensus: Adjusting advice following competency-based resistance in L2 writing tutorials at a British University. *Classroom Discourse* 9: 267–87. [\[CrossRef\]](#)
- Licoppe, Christian, and Julien Morel. 2012. Video-in-Interaction: “Talking Heads” and the Multimodal Organization of Mobile and Skype Video Calls. *Research on Language and Social Interaction* 45: 399–429. [\[CrossRef\]](#)
- Luff, Paul, Christian Heath, Hideaki Kuzuoka, Jon Hindmarsh, Keiichi Yamazaki, and Shinya Oyama. 2003. Fractured ecologies: Creating environments for collaboration. *Human-Computer Interaction* 18: 51–84. [\[CrossRef\]](#)

- Malabarba, Taiane. 2015. O percurso do agir interacional no trabalho docente: Do projeto de ensino às participações contingentes em sala de aula de língua inglesa [The Interactional Trajectories of English-as-a-Foreign-Language Teaching: From the Instructional Project to Contingent Participations]. Unpublished Ph.D. dissertation, Universidade do Vale do Rio dos Sinos, São Leopoldo, Brazil.
- Mondada, Lorenza. 2007. Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers. *Discourse Studies* 9: 194–225. [CrossRef]
- Mondada, Lorenza. 2014. The local constitution of multimodal resources for social interaction. *Journal of Pragmatics* 65: 137–56. [CrossRef]
- Mondada, Lorenza. 2015. Ouverture et préouverture des réunions visiophoniques. *Reseaux* 194: 39–84. [CrossRef]
- Mondada, Lorenza. 2019. Conventions for Multimodal Transcription. Available online: <https://www.lorenzamondada.net/multimodal-transcription> (accessed on 31 October 2021).
- Moorhouse, Benjamin Luke, Yanna Li, and Steve Walsh. 2021. E-classroom interactional competencies: Mediating and assisting language learning during synchronous online lessons. *RELC Journal*, 1–15. [CrossRef]
- Mortensen, Kristian. 2009. Establishing reciprocity in pre-beginning position in the second language classroom. *Discourse Processes* 46: 491–515. [CrossRef]
- Nguyen, Hanh, Ann Tai Choe, and Cristiane Vicentini. 2022. Opportunities for second language learning in online search sequences during a computer-mediated tutoring session. *Classroom Discourse*, 1–19. [CrossRef]
- Olbertz-Siitonen, Margarethe. 2015. Transmission delay in technology-mediated interaction at work. *Psychology Journal* 13: 203–34. Available online: <https://jyx.jyu.fi/handle/123456789/51325> (accessed on 31 October 2021).
- Oloff, Florence. 2013. Embodied withdrawal after overlap resolution. *Journal of Pragmatics* 46: 139–56. [CrossRef]
- Oloff, Florence. 2019. Okay as a neutral acceptance token in German conversation. *Lexique* 25: 197–225. Available online: https://lexique.univ-lille.fr/data/images/numero-25/10_Oloff_ok%20pagin%C3%A9_FINALE_20-12.pdf (accessed on 31 October 2021).
- Park, Innhwa. 2017. Questioning as advice resistance: Writing tutorial interactions with L2 writers. *Classroom Discourse* 8: 253–70. [CrossRef]
- Ruhleder, Karen, and Brigitte Jordan. 2001. Co-Constructing Non-Mutual Realities: Delay-Generated Trouble in Distributed Interaction. *Computer Supported Cooperative Work (CSCW)* 10: 113–38. [CrossRef]
- Rusk, Fredrik, and Michaela Pörn. 2019. Delay in L2 interaction in video-mediated environments in the context of virtual tandem language learning. *Linguistics and Education* 50: 56–70. [CrossRef]
- Sacks, Harvey. 1992. *Lectures on Conversation*. Oxford: Basil Blackwell, 2 vols.
- Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50: 696–735. [CrossRef]
- Schegloff, Emanuel A. 1997. Whose text? Whose context? *Discourse and Society* 8: 165–87. [CrossRef]
- Schegloff, Emanuel A. 2000. Overlapping talk and the organization of turn-taking for conversation. *Language in Society* 29: 1–63. [CrossRef]
- Schegloff, Emanuel A. 2007. *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge: Cambridge University Press, vol. 1.
- Schegloff, Emanuel A. 2010. Some Other “Uh(m)”s. *Discourse Processes* 47: 130–74. [CrossRef]
- Schneider, Jan Georg. 2017. Medien als Verfahren der Zeichenprozessierung: Grundsätzliche Überlegungen zum Medienbegriff und ihre Relevanz für die Gesprächsforschung. *Gesprächsforschung—Online-Zeitschrift Zur Verbalen Interaktion* 18: 34–55.
- Sert, Olcay. 2017. Creating opportunities for L2 learning in a prediction activity. *System* 70: 14–25. [CrossRef]
- Sert, Olcay. 2021. Transforming CA Findings into Future L2 Teaching Practices: Challenges and Prospects for Teacher Education. In *Classroom-Based Conversation Analytic Research: Theoretical and Applied Perspectives on Pedagogy*. Edited by Silvia Kunitz, Numa Markee and Olcay Sert. New York: Springer International Publishing AG, pp. 259–79.
- Seuren, Lucas M., Joseph Wherton, Trisha Greenhalgh, and Sara E. Shaw. 2021. Whose turn is it anyway? Latency and the organization of turn-taking in video-mediated interaction. *Journal of Pragmatics* 172: 63–78. [CrossRef]
- Stevanovic, Melisa. 2018. Social deontics: A nano-level approach to human power play. *Journal for the Theory of Social Behavior* 48: 369–89. [CrossRef]
- Svennevig, Jan. 2000. *Getting Acquainted in Conversation: A Study of Initial Interactions*. Amsterdam: John Benjamins Publishing.
- Uskokovic, Budimka, and Carmen Talehgani-Nikazm. 2022. Talk and Embodied Conduct in Word Searches in Video-Mediated Interactions. *Social Interaction. Video-Based Studies of Human Sociality* 5: 1. [CrossRef]
- Walsh, Steve. 2006. *Investigating Classroom Discourse*. New York: Routledge.
- Waring, Hansun Zhang. 2005. Peer tutoring in a graduate writing centre: Identity, expertise, and advice resisting. *Applied Linguistics* 26: 141–68. [CrossRef]
- Waring, Hansun Zhang. 2011. Learner Initiatives and Learning Opportunities in the Language Classroom. *Classroom Discourse* 2: 201–18. [CrossRef]
- Waring, Hansun Zhang. 2016. *Theorizing Pedagogical Interaction: Insights from Conversation Analysis*. New York: Routledge.
- Wigham, Ciara R. 2017. A multimodal analysis of lexical explanation sequences in webconferencing-supported language teaching. *Language Learning in Higher Education* 7: 81–108. [CrossRef]