

Social Interaction. Video-Based Studies of Human Sociality.  
2022 Vol. 5, Issue 3  
ISBN: 2446-3620  
DOI: 10.7146/si.v5i3.131029

## ***Social Interaction***

### ***Video-Based Studies of Human Sociality***

---

### **The Regarded Listener: Transcribing the Multiparty Social Ecology of Children's Collaborative Storytelling**

---

***Sarah Jean Johnson<sup>1</sup> & Frederick Erickson<sup>2</sup>***

<sup>1</sup>University of Texas at El Paso

<sup>2</sup>University of California, Los Angeles

#### *Abstract*

*Current theories of human social interaction characterize interactants, both speakers and hearers, as possessing a rich cognitive and reflexive life achieved through collaboratively built action. However, prevailing transcription approaches, which tend to neglect listening activity, do not adequately incorporate such understandings of the phenomena they represent transcriptively. We review the history and scholarship of video-based studies of human sociality. We then present an example of horizontal transcription to demonstrate the utility of this approach in capturing the spatial, temporal, and visual components of human social interaction—in this case, that of young children telling stories as they are writing them.*

**Keywords:** multimodal transcription, participation frameworks, oral narrative, peer interaction

## 1. Introduction

On a school day in December, Perry, with four of her kindergarten classmates, are seated around a table, talking and writing. They are generating ideas about brief stories they will compose about happenings in their families at Christmas time. How might we account descriptively and analytically for what they are doing as they listen and speak with one another? What approaches to transcription might shed light on how they make meaning from moment to moment?

In a now classic statement titled “Transcription as theory,” Ochs (1979) maintained that the real-time conduct of interactional behavior is so complex and multidimensional such that any attempt to transcribe it from machine recordings is inevitably partial and selective. Driving such selection are theoretical assumptions, both explicit and implicit, concerning the nature of the phenomena that are being described. What is foregrounded and backgrounded in the transcript—which aspects of speech, which aspects of non-vocal activity, which participants are being featured as central or peripheral in the interaction—result from research interests and commitments that are ultimately ontological—assumptions about the fundamental nature of co-present interaction itself. (Currently, with two successive generations of experience with such transcribing, we can now see that transcription is not only guided by theory, but it can also inform revisions in theory—as limitations in transcription approaches become apparent, awareness of these limitations can stimulate theory development.)

Little more than a decade before the statement by Ochs, Goffman (1964) had characterized the relations among participants in interactional encounters as grounded in patterns of mutually influencing attention and action. He defined the social situation as “an environment of mutual monitoring possibilities” and defined encounters as “an ecological huddle wherein participants orient to one another and away from those who are present in the situation but not officially in the encounter” (p. 135). These definitions presume ontologically that the organization and conduct of social interaction in real time is fundamentally ecological—the maintenance of continuous relationships of mutual influence among participants. (This ecological perspective on the local conduct of social action face to face is consonant with Weber’s basic definition of social action more generally: “Action is social, in so far as by virtue of the subjective meaning attached to it by the acting individual it takes account of the behavior of others and is thereby oriented in its course” (Weber, 1922/1978, p. 4). (Admittedly, in that statement Weber was referring to economic activity in society more broadly, as in commerce or in a stock market, but taking account of the actions of others is also fundamental at the micro-social level of the face-to-face encounter.)

Goffman's definitions emphasize co-ordination of attention and monitoring among interactional participants. He invoked the metaphor of the American football huddle, in which players on the field confer about immediate next moves in the game by forming a circle, with each participant leaning forward watching the others while listening to the speech of a focal participant in the scene, usually the quarterback. In developing this ecological perspective on the organization and conduct of social interaction, Goffman was influenced quite early in his career by firsthand acquaintance in the California Bay Area with researchers who were trying out the first attempts at systematic transcription and analysis of verbal and nonverbal behavior together, using sound cinema film. These pioneers of the mid-1950s included Jürgen Reusch, Gregory Bateson, and others in a research group at the Veterans Hospital in Palo Alto, and an interdisciplinary research group which included Bateson at the Center for Advanced Study in the Behavioral Sciences, who initiated there a project that came to be called "The Natural History of an Interview." (For a discussion, see Leeds-Hurwitz, 1987 and Kendon, 1990). By the mid-1960s for a next generation of scholars, this approach was coming to be called "context analysis" (Scheflen, 1973). One of its practitioners, William Condon, emphasized the importance of "whiles" in the social ecology of interaction: while participant A is doing X, participant B is doing Y (Condon & Ogston, 1967). This insight points to the importance of transcribing to take account of simultaneity in the co-ordination of actions as well as of sequential organization of action. Emphasis on the latter has been the hallmark of what emerged in the 1970s as "conversation analysis," with an approach to the transcription of speech that foregrounds its sequential aspects (Sacks et al., 1974). The approach of context analysis, in contrast, was to attempt a transcription that took account of both simultaneity and sequence in the conduct of interaction, a harbinger of the recent "embodiment turn" in studies of interaction and concern for "co-operative action," developments that will be discussed further below.

When people interact together, they don't do it with their eyes closed, as if listening to one another through a keyhole or on the telephone. With eyes open they are literally "face to face" in copresence, using the entire sensorium—looking, listening, smelling, touching, experiencing heat, employing kinesthetic perception—in monitoring one another's actions so that each participant is able to take account of the simultaneous and sequential actions of others while that participant is also taking actions themselves. It is not that some participants in the encounter are active at some moments while others are passive. Everyone is doing something; no one is ever doing nothing. Listeners are influencing speakers as speakers are influencing listeners. As McDermott (1976) put it, "people in interaction constitute environments for one another" (p. 36). And as Erickson (1986, p. 295) has observed, engaging in social interaction with others is like climbing a tree that is climbing you back at the same time. (See also the discussion in Erickson, 2015.)

In sum, when we combine the perspectives of Goffman and context analysis on the ecological nature of interaction with the insights of Ochs concerning transcriptive representation, it is apparent that there is a need for approaches to transcription that foreground the mutuality of monitoring and influence among all interactional participants. One way to accomplish this is to transcribe speech and non-verbal activity horizontally along a constant timeline rather than by means of playscript representation. What follows in this article is further discussion of literature on the way toward the current “embodied turn.” An extensive example of horizontal transcription is then presented, with explanation of its transcription conventions and discussion of the interaction that the transcription makes visible. The article concludes with a brief discussion of the implications of this approach for video-based research on face-to-face interaction.

## 2. Survey of Horizontal Transcription

An early innovation in the graphic portrayal of simultaneous action in time and space comes from dance instruction books of the late Renaissance and Baroque eras (see, for example, the illustrative plates in Arbeau’s *Orchésographie* published in Langres in 1589, with a modern English translation published in 1967, and plates from Tomlinson’s *The Art of Dance* published in London in 1735 and reprinted in Tufte (1990, pp. 114-115)). In the 20<sup>th</sup> century horizontal transcription for documenting and analyzing human movement is credited to Rudolph von Laban, who described his complex notational system for recording dance, coined “Labanotation,” in *Schrifttanz (Written Dance)* in 1928.<sup>1</sup>

As noted in the introduction, in the tradition of interactional studies, early pioneers in the United States who independently developed horizontal transcription methods for analyzing the synchrony between talk and movement in multiparty interaction were the Natural History of an Interview Group, Condon (Condon & Ogston, 1966), Scheflen, (1964), Kendon (1970), and Erickson (Erickson & Shultz, 1982). While each analyst developed different representational conventions for semiotic modes, the template for their transcriptions followed a similar format in which the video frame number was indicated on the “x axis” (the standard rate for film was 24 frames per second) forming a timeline, and the participants and their verbal and nonverbal modes of communication were overlaid vertically on the “y axis.” This structure is the basis of contemporary *partitur* (German for “musical score”) transcription software and for the transcription method presented in this manuscript.<sup>2</sup> Erickson, who was

<sup>1</sup> See also, Klein (2020) for contemporary use of *partitur* transcription software to graphically represent and preserve dance choreography.

<sup>2</sup> Konrad Ehlich and Jochen Rehbein, who work within the field of functional pragmatics, are German contemporaries of the “context analysis” descendants of the original Natural History of an Interview Group. Ehlich and Rehbein’s score notational system (see Ehlich & Rehbein, 1976),

trained as a musician and ethnomusicologist, adapted this method for what he termed “quasi-musical transcription,” i.e., notations that display speech and movement rhythms through the use of musical notes within a metered measure (Erickson, 1982, 2009; see also Scollon, 1982). By demonstrating how context is an enacted and mutually built social environment, these early studies countered the dominant view at the time of context being separate from and inconsequential to verbal behavior.

Another important early pioneer in the development of transcription methods that capture simultaneous embodied interaction is C. Goodwin (1981). Working within both a context analysis and a CA approach, Goodwin proposed a notational system for depicting the trajectory of gaze in which the auditor’s continued gaze is shown by a horizontal line beneath the speaker’s talk, with keyboard symbols (e.g., periods, commas) indicating the beginning and retraction of the gaze. In the following decades, his transcription method for analyzing the mutual coordination of the body and speech in multiparty interaction has been widely adapted by those working within a CA framework. The reader, for example, can refer to Alkemeyer, Brümmer, and Pille (2017) for the unfolding of gesture in coordination with talk; to Streeck (1994) and C. Goodwin and M. Goodwin (1986) for the trajectory of gesture and gaze; and to Churchill (2022) for the coupling of gaze, gesture, and piano playing in the education of attention in musical enskillment.

The field of CA, self-described as the study of “talk in social interaction,” had initially placed central emphasis on the organization and conduct of speaking, with non-vocal activity of listeners either not considered at all (as in the transcription and analysis of telephone conversations), or considered peripherally, as secondary to talk. However, the field of CA experienced an “embodied turn” at the beginning of the current century and, as did the early pioneers, began to exploit the visual affordances offered by video recording technology and develop transcription approaches for representing the body. A transcription approach created by Mondada (2018) through the past two decades, which builds on these earlier traditions, has achieved standardized use in the field. Mondada’s technique employs CA conventions in its use of keyboard symbols for phenomena within the vocal stream (e.g., colons for lengthened speech) while expanding the approach in its additional application of keyboard symbols for embodied actions. Aspects of synchronization

---

termed HIAT (*Halbinterpretative Arbeitstranskriptionen*), is an early prominent example of horizontal transcription and has been highly influential among discourse-based research in Europe (see Ehlich & Rehbein, 1976)—a tradition from which partitir-based transcription software developed (as discussed further in Ehlich, 2014). The HIAT method employs the x and y axis within a score transcript to display the temporal and synchronous features of multiparty verbal and nonverbal communication with its own codified use of symbols (e.g., rising and falling dots above speech to denote intonation) and abbreviations (e.g., different body parts are specified with two-letter abbreviations).

of talk and embodied conduct of multiple participants is achieved by the inclusion of multiple tiers below the speaker that are dedicated to nonverbal behavior. The text format remains vertical on the page with signs, such as lines, arrows, and symbols, indicating the delimitation of the beginning and end of an embodied action. Images are presented in the manuscript with a symbol in the transcript indicating the location for the image. A very fine degree of granularity in representation of social interaction is possible with this technique, the convention of which can be found on Mondada's website (Mondada, 2021). A sampling of recent research shows the broad applicability of the technique to interactional studies. It has been used to study bodily formations in kissing (Mondada et al., 2021), expressions of gratitude as part of passing objects (de Souza et al., 2021), the use of embodied resources in directing students' attention to a pedagogical task (Routarinne et al., 2020), the suspension of manual activity and body torque within the organization of conversational repair (Kamunen, 2019), and professional touch in speech and language therapy for aphasia patients (Merlino, 2021).

Within the CA tradition, transcription methods developed by M. Goodwin and C. Goodwin arguably provide the clearest highlighting of phenomena of analytic interest. For example, C. Goodwin (2018) builds an argument about the relevance of the body and structures in the environment to emerging talk in his analysis of co-operative action in girls' hopscotch. Goodwin's transcripts incorporate the hopscotch grid and a drawing of the participants' configuration within the grid. Graphic images of the girls' foot placement within the hopscotch grid and hand gestures are placed above the talk to demonstrate the importance of the timing of gestures and talk and of the geography of foot placement to the multi-modal performance of the game. M. Goodwin and Cekaite (2018; see also, M. Goodwin, 2021) in their study of "intertwinings," or hugs within family interaction, present multiple frame grabs from video horizontally across the page with vocalization and pitch tracks below the images. By looking at how speech production is tied to embodied configurations and the evolution and trajectory of hugs, they build arguments around how intimacy is an achieved accomplishment. Others studying the role of touch in the accomplishment of human social interaction have taken a similar transcription approach (Heath & Luff, 2021; Iwasaki et al., 2019; Meyer & Wedelstaedt, 2020).

Another framework for multimodal transcription is provided by Norris (2004), who builds her analytic model from the traditions of mediated discourse analysis (Scollon, 2001), interactional sociolinguistics (Gumperz, 1982), and multimodality (Kress & Van Leeuwen, 2001). Norris' approach relies in a large part on annotated video stills as a way to reflect how nonverbal modes, artifacts, and the material world figure into human meaning-making. Her transcriptions, however, tend to give less emphasis to the timing and mutual coordination of ongoing action. A graphic approach which highlights sequential and simultaneous structure of interaction is provided by Lerner

et al. (2011), who, using a series of annotated images, demonstrate how two prelingual infants make sense of the activity of a caregivers' feeding in structuring their own embodied participation. Iwasaki et al. (2019) provide an example of a similar approach in their analysis of tactile signed interaction.

Researchers are also increasingly using comic book format methods—video or drawn stills of images with embedded speech bubbles which are arranged to map temporal dimensions—as a way to transcribe embodied moment-to-moment interaction (Laurier, 2014; Plowman & Stephen, 2008). Plowman and Stephen (2008) argue the format is ideal “for representing a complex process within a sequence of static images, making it easier to isolate key gestures and responses, nonverbal means of interaction and ways in which multiple participants orient to each other and features of the environment” (p. 560).

In our survey of existing research, we identified few recent forms of horizontal transcription that are not hybridized with vertical methods or reliant primarily on graphic stills of images. An exception is Bezemer and colleagues' study of professional activity in the “operating theater,” which draws upon CA and ethnomethodological frameworks. These scholars apply horizontal transcription methods developed by Heath (1986) in their argument that the surgeons' technical skills, such as cutting sutures, cannot be viewed separately from their communication skills with the scrub nurse in terms of understanding the accomplishment of surgical operations.<sup>3</sup>

A second recent example of horizontal transcription is provided by İkizoğlu (2019, 2021), whose dissertation examined video recordings of multilingual family interaction that was facilitated by a voice translation application on a mobile phone. İkizoğlu uses the output option built into ELAN software for embedding the horizontal transcript in her manuscript. The transcript provides a designated line with two tiers for the phone—showing the sounds it makes and its physical location. By showing how the family orients to the application—for example, through their gaze and laughter—she argues the technology is a participant in the family's collaborative meaning making and, hence, family dynamics and relations.

The transcription approaches we survey reflect tremendous advancement from the traditional playscript transcription approach. They complicate and expand our understanding of collaborative meaning-making as part of human interaction by not privileging a single mode of action as the object of study or experience (Sicoli, 2018). There remain significant limitations in their utility, however. For one, there is a

---

<sup>3</sup> Bezemer is affiliated with a research group housed at the University of London which is dedicated to research and training on multimodality and offers a “transcription bank” for researchers to reflect on and compare different forms of transcription (MODE Transcription Bank, 2021).

tendency within the recent approaches to rely on the typeset of the keyboard for symbolic representation, word processing software for document creation, and image stills for displaying nonverbal modes. The plethora of symbols and series of images (that rarely have phenomena of interest highlighted) pose challenges for readability and interpretation even for those trained in a particular transcription technique. Secondly, the vertical format of the typeset page is constraining for the presentation of data and for representing its temporal dimensions. Graphic design software and the ability of online journals to hyperlink to documents, videos, and images open up new vistas for representation that could be further exploited in transcriptions. This journal, for example, provides an innovative approach in its capability to embed within transcripts GIFS (animated graphic interchange formats).

We acknowledge that the horizontal transcript we present in this manuscript does not altogether solve the issues we allude to and may, in fact, pose additional challenges to readability due to our novel use of transcription conventions. We developed the transcript, however, in the spirit of a “discovery procedure”—that is, an exploration of how a departure from established transcription conventions can help open our professional vision to unexpected phenomena (Duranti, 2006, p. 307; see also, Slembrouck, 2007). In this manuscript, we draw from both our reflexive stance towards our transcription procedures and our analysis to argue for innovations in the entextualization of video-based discourse in a manner that honors the simultaneous and sequential nature of conjoint social action (Bucholtz, 2000).

### **3. Description of Transcription**

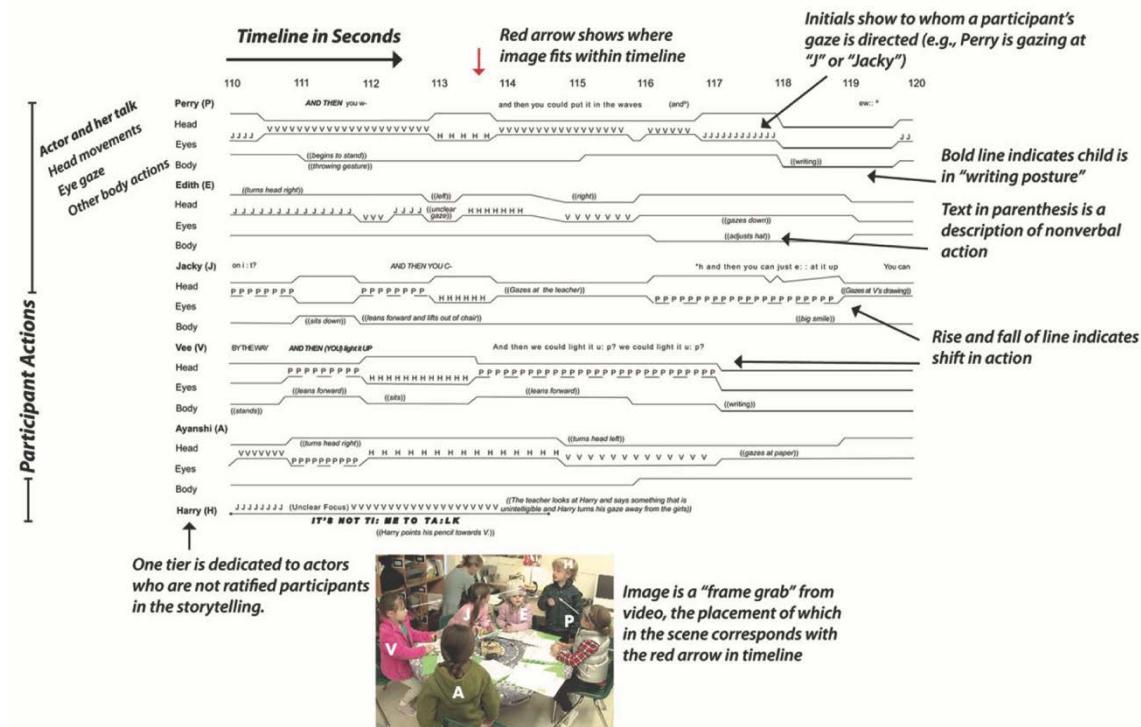
The first author was a doctoral student of Frederick Erickson (the second author) as well as of Charles and Marjorie Goodwin at the University of California, Los Angeles. She adapted the transcription approach presented in this manuscript from methods she was taught from these pioneers in the field. The transcript displays the various actions of participants in relation to a shared timeline which shows the “real time” of the event being transcribed from video. The timeline is displayed horizontally across the top of the page and the actions of the participants are overlaid vertically across the timeline in tiers. Figure 1 provides a graphic illustration of the conventions described in this section.

The five child participants—Perry, Edith, Jacky, Vee, and Ayanshi—are each given a horizontal space within the transcript. Within each participant’s space, there are four horizontal tiers: the top one for that person’s speech when it occurs, the next lower tier for her head movement (e.g., nods), the tier below that for her eye movement (e.g., gaze direction), the tier below that for her body movement (e.g., shifts in postural position, reaching for an object) and for any other action that seems salient. The line extending horizontally across at each tier shows the continuation of

some action and the rise and fall of the line indicates a shift of action. Lines which zig-zag indicate rapid movement, such as nodding the head repeatedly. We include brief textual descriptions of the action in the transcript above the line to which they correspond.

The “eyes” tier indicates the onset and offset of gaze that is directed by that person to another of the participants. The identity of the participant to which the gaze is directed is indicated by successive initials of that person’s name. The string of successive initials lasts as long as the gaze is sustained. Thus, gaze directed at Perry is indicated by “P P P P P P P P P,” while gaze directed at Jacky is indicated by “J J J J J J J.”

**Figure 1.**



The transcript additionally employs conversation analysis conventions developed by Gail Jefferson and described in Sacks et al. (1974, pp. 731-733) for transcription of the vocal stream. Deviations from Jefferson's conventions are the following: (1) capitalization of text indicates increased volume and bold italics show some kind of vocal emphasis (e.g., changes in pitch); (2) word-by-word speech is linked in approximation with the timeline so that the text of more rapid speech is condensed and that of slower talk is spread out, thus allowing for the representation of the

synchrony of overlapped speech. For this stage of the transcription process, we used the audio software, *Audacity*, which provides a timeline and prosody visualization which is linked to the uploaded audio file.

The initial version of this transcript was drafted on a 34-inch-tall graph paper roll with half-inch grid patterns. The completed transcript, which represents 130 seconds of interaction, extends about 25 feet. The transcript excerpts we present in this manuscript are a digital version of this original.

We chose the analogue approach to transcription due to our lack of satisfaction with options for exporting transcripts offered by video annotation software. For example, a transcript from a widely used open-source software, ELAN, presents descriptive text for all multimodal actions on a straight line for each horizontal tier (essentially in the format of a Microsoft Excel document); this uniformity and textual density creates a challenge for reading and interpretation. A second popular annotation software, Feldpartitur, offers many visual symbols, such as smiley faces, hand gestures, and arrows, for use along with textual description. The symbols, however, are generally not useful in depicting interactional phenomena. In our survey of horizontal transcription (discussed previously in this article), we identified few studies that used exported transcripts from software programs, which is likely due to this limited flexibility in terms of output format. Conducting a first-stage transcription with the software is a viable option that we acknowledge might have less potential for inaccuracies. By producing an analog transcript, however, we were able to design the transcript in a manner that could easily be copied and digitized by a graphic designer.

#### **4. Data**

The data for the transcription derive from a video recording of a kindergarten classroom at a progressive private school in Southern California. The class was led by two veteran teachers who follow the Writer's Workshop curricular approach developed by Lucy Calkins (2001). The curriculum provides extended time for children to write independently while the teacher holds writing conferences with individuals or small groups of children. The first author conducted field work in this classroom for her dissertation on peer sociality and literacy learning in the academic calendar year of 2011 to 2012 (Johnson, 2015).

Prior to reading the following narrative description of the scene, it will be helpful for the reader to watch the accompanying video of the scene.

>>video<<

## 5. Narrative Description of the Storytelling Scene

It is shortly before the winter holiday, and the girls (Perry, Edith, Jacky, Vee, and Ayanshi) are sitting at a round table writing holiday-related stories. Some are in the “publishing phase” of the writing process and are frequently reaching for markers located in the center of the table as they color the pictures illustrating their stories. Perry is still writing her story and is holding a pencil in her hand. The children are dressed for a chilly Southern California winter. One wears a puffy ski vest, another girl dons a knit hat, and three are wearing soft zip-up hoodies.

The journalism center is located behind the table where the girls are writing, and the teacher is holding a writing conference in this location with a boy named Harry. The young boy eavesdrops on the girls as the teacher reads his assignment.

As the children are writing stories, they are also telling stories. In our selected scene, these interconnected activities have multiple constituent parts which we infer based on conversational narrative structure (Polanyi, 1982) as well as the interactional patterns of participants through which they indicate a shift in participation framework (e.g., changes in posture, speech, and/or prosody). We detail how our transcript makes clear these participation frameworks later in this manuscript. We present a gloss of the primary constituent parts of the scene presently as a means of orienting the reader to the interaction our transcript depicts. The parts include: introduction of the topic, narrative setup, traditional narrative, hypothetical narrative, and closing.

### 5.1 Introduction of the topic

The scene begins with children writing quietly in the company of one another. Perry is humming and Jacky is singing quietly. Vee, while still writing, announces to her friends she got her Christmas tree. Other children “piggyback” on this topical substrate by also announcing their families’ acquisition of Christmas trees or, in the case of Ayanshi, that she didn’t get her tree yet.

### 5.2 Narrative setup

As the children report on their Christmas trees, Perry, sitting tall as if to command attention, opens up a second conversational floor to tell her friends about the Christmas she is writing about. Her talk overlaps that of her peers’ Christmas tree announcements; at the same time, some children are engrossed in their writing. As a result, her narrative preface is extended as she seeks the shared attention of her peers (Ochs & Capps, 2001).

### 5.3 Traditional narrative

Upon achieving the gaze of all her friends, Perry begins to tell a traditional narrative as defined by Labov (1972, pp. 359-360) as “recapitulating past experience by matching a verbal sequence of clauses to the sequence of events which (it is inferred) actually occurred.” Her brief story is about Santa’s visit to her home and his shaking off snow from his suit, which created a great mess.

Perry’s narrative is interrupted as her peers, Jacky and Vee, discuss the palatability of snow—theirs is a positive assessment to which Perry disagrees. Perry continues her narrative by talking about how the snow was too melted for a snow fight. This shift of her story to the hypothetical realm (what was not possible but might have happened had the snow been firmer) provides a prelude to the next section of the storytelling.

### 5.4 Hypothetical narrative

The transition to the hypothetical narrative—a narrative about something that never happened in the past but might, nevertheless, occur at some time—corresponds, in this case, with a grammatical turn to the conditional tense. This narrative emerges as a “second story” (Sacks, 1967) to Perry’s story of the snow falling off Santa’s suit with Jacky’s statement, “You know, if you put a snowball in- in your freezer, it will not melt. It will turn into ice.” Because the story is not based on past experience that might be privileged to a sole narrator, the story is now open to multiple narrators’ participation. Notably, the story has also been established as “tellable” from the initial narrative structure provided by Perry and Jacky. As such, less competent storytellers have an entry point without providing justification for their rendering as long as it continues in the same format as the emerging storyline. This is what happens as multiple girls chime in through their use of “what if” and “if then” clauses to improvise the conditional possibilities of what might occur if you put a snowball in the freezer. Talk is overlapping as the children compete by offering twists to the narrative and evaluate others’ contributions. Perry is able to continue to dominate the storytelling and tells the resolution of the story, which has to do with surfing on the ice. Ayanshi, who has been writing while also attending to the developing story, is able to inject a “but” into the competitive stream of talk, indicating her intended, yet unfulfilled, proposition for the story to take an alternate trajectory.

### 5.5 Closing

The scene closes as children resume writing. Jacky shuffles papers in her writing folder while singing, “I love Perry...” in a repeat refrain—a song which appears to

acknowledge Perry's hierarchical position gained from her display of skill in the storytelling.

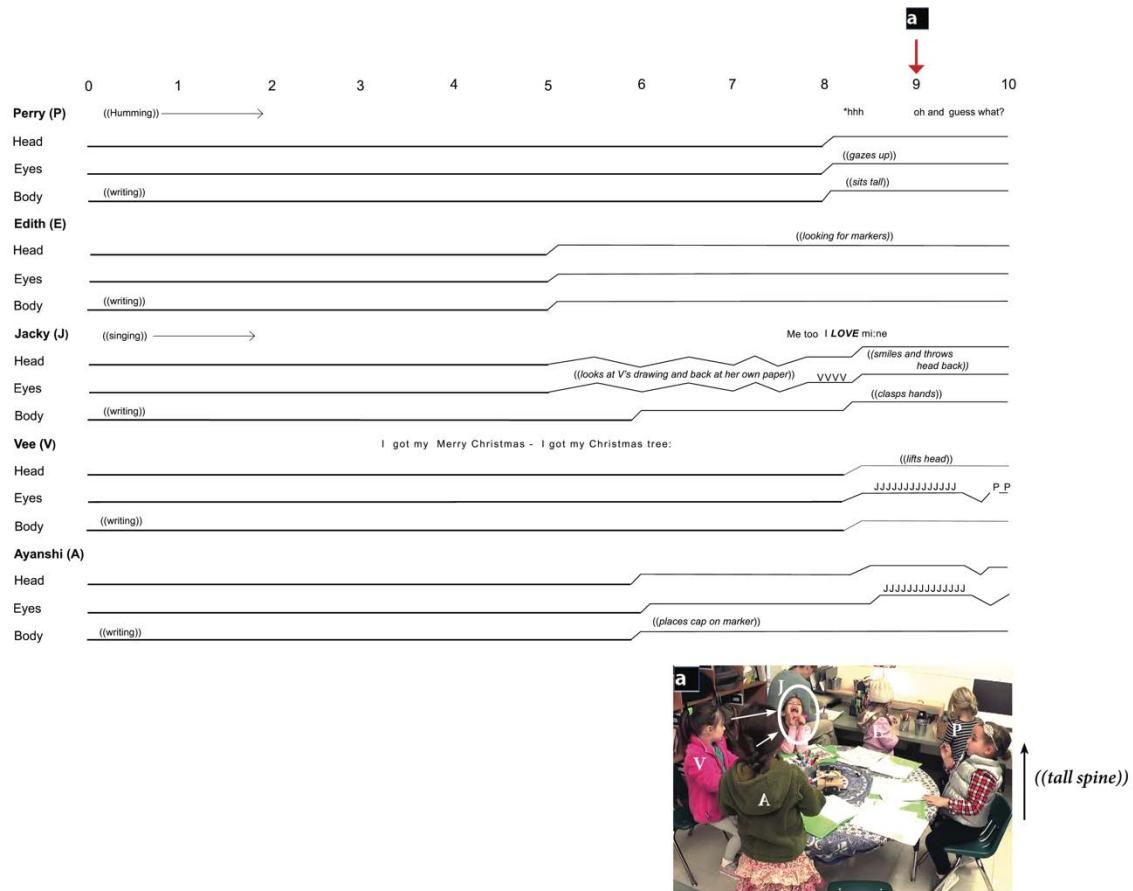
## **6. Examining Children's Storytelling as Represented through Horizontal Transcription**

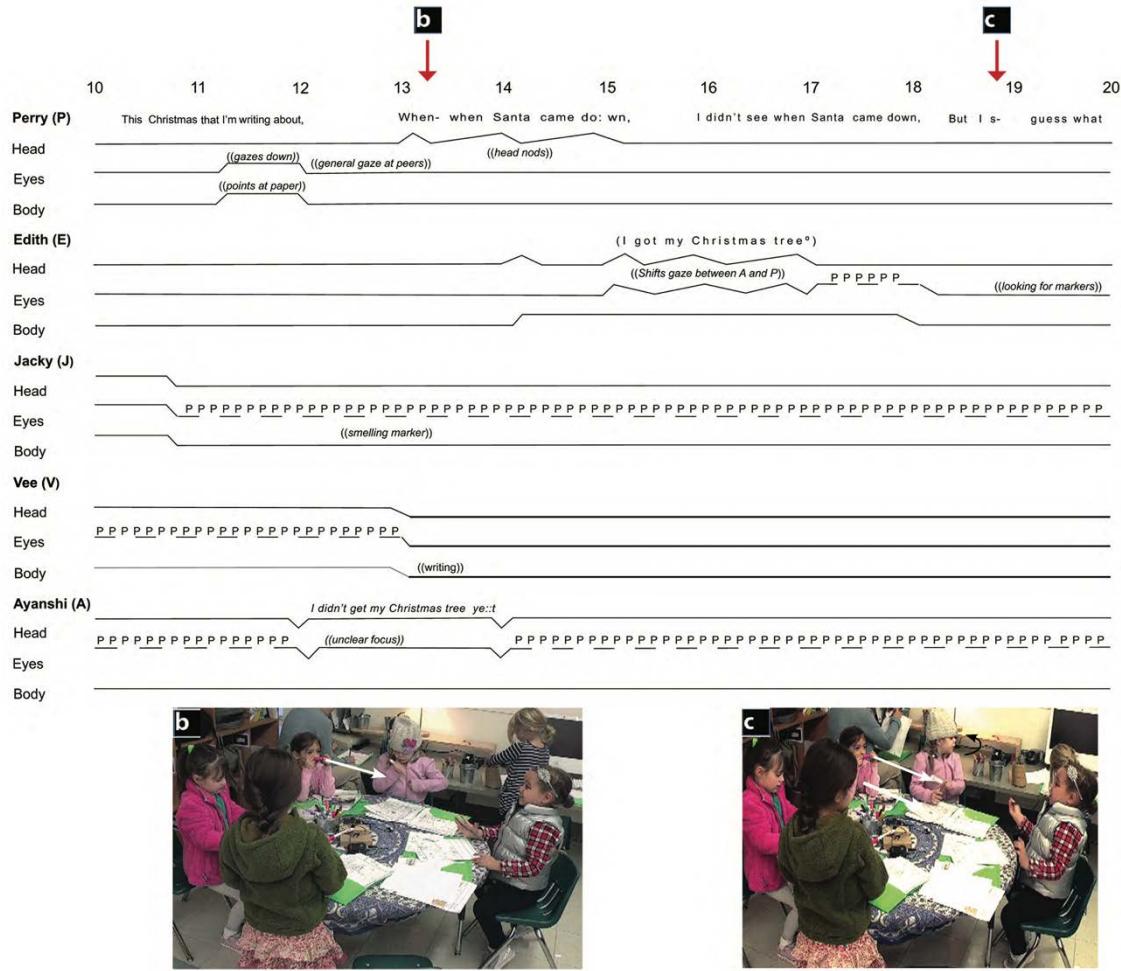
In the sections which follow, we take a closer examination of the storytelling scene using our transcription. In the first section, we visualize the audience anew as differentiated and agentive (as opposed to uniform and passive). The second section looks at the diversity of resources which both the speaker and audience bring to bear on their story production. The third and fourth sections look at the consequential nature of children's activity in terms of achieving cooperative action. The final section examines how children build their lifeworld around storytelling through the enactment of complex participation frameworks where children assume a variety of roles within the interaction. In each section, we guide the reader to refer to the transcript so as to see our representation of these participation frameworks.

### **6.1 Launching a story and the differential agency of the audience**

The launching of a narrative is not achieved by an individual speaker but instead is dependent on the cooperation of all interlocutors (Ochs & Capps, 2001). This point is made visible by our transcription which illustrates the complex activity occurring as Perry attempts to launch her story about Santa coming down the chimney and shaking the snow from his suit.

**Figure 2.**



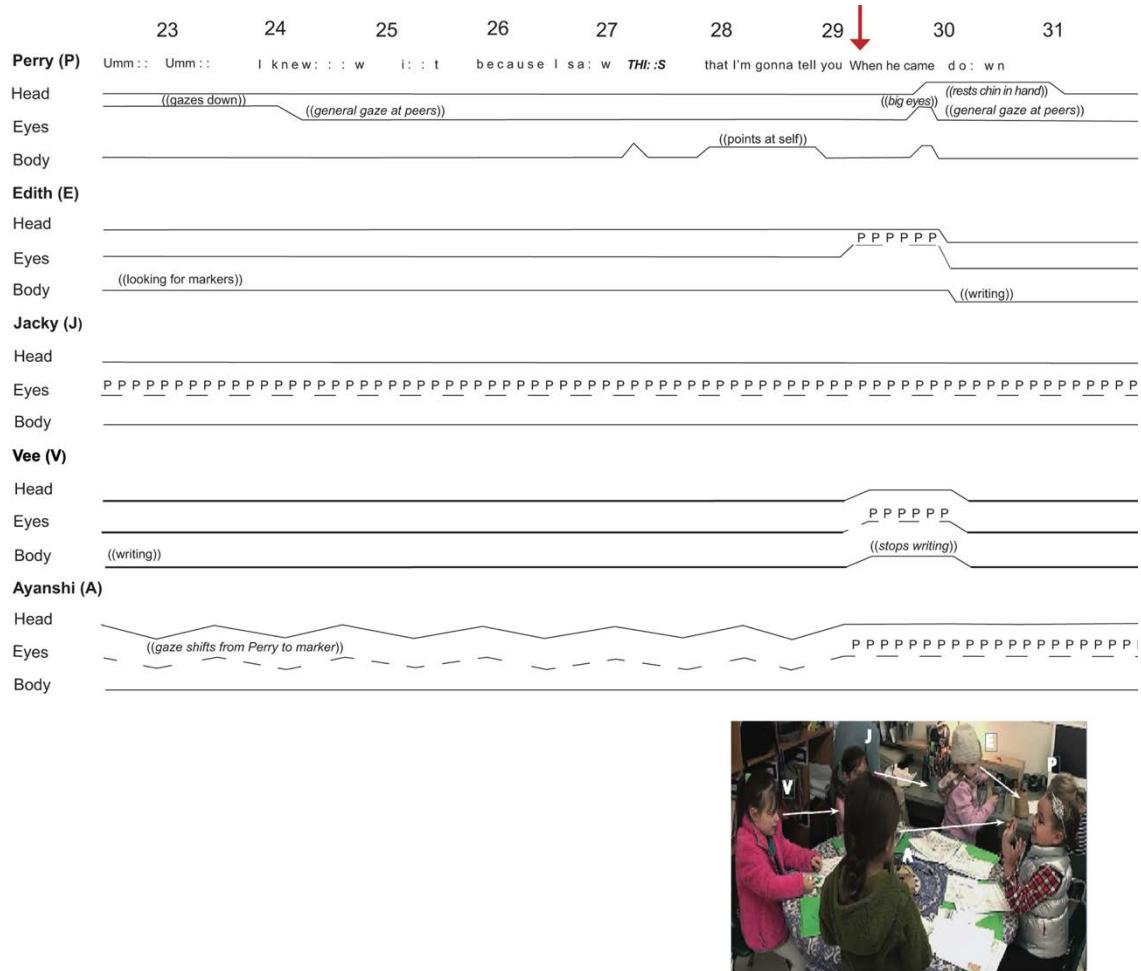


Our transcript illustrates the two parallel and temporally unfolding conversational floors and the impact of this multiple activity on Perry's talk. Children on one conversational floor are announcing the acquisition of Christmas trees. This progresses in so-called normal turn-taking fashion. Let us examine this in Figure 2. Vee states, "I got my Merry Christmas- I got my Christmas tree" (3 seconds) and Jacky "piggybacks" on her friend's announcement, saying "me too, I love mine." Her statement is coupled with an embodied performance of smiling and throwing her head back (7.5 seconds). Perry, however, opens a second conversational floor overlapping Jacky's exuberant exclamation. Perry suddenly sits up tall as if to command attention, inhales sharply and says "oh and guess what", and continues, "this Christmas that I'm writing about" (9-12 seconds). She does not have a unified audience, however, which would allow her to control the floor for an extended time to tell a story. Ayansi and Edith both time their interjections about Christmas trees during two different brief pauses in Perry's talk, though Perry continues talking over their contributions (12-18 seconds).

This complex interactional environment, where there are two conversational floors overlaying the activity of writing is evident not only through the talk. Notice in the transcript in Figure 2 the fracturing of children's gaze, for example. (As a reminder of transcript conventions, the children's initials (e.g., "J" and "V") indicate a participant is gazing towards them.) Jacky and Vee are able to capture the gaze of some peers with their announcements (3-9 seconds). As Perry is talking, her peers' gazes are divided between looking at Perry and looking at their writing. Edith's gaze briefly shifts between the prior speakers on both conversational floors—Perry and Ayanshi—as she quietly announces that she also got her tree (15 seconds).

This multiplicity of ongoing and simultaneous activity does not occur in a separate ecosystem as Perry's production of talk but instead is mutually influential. This point is evident in Perry's multiple pauses, restarts and "um" tokens. These interactional stumbles are not due to her innate inability to articulate the story but rather to the complex interactional ecosystem in which her story emerges (C. Goodwin, 1981). Evidence for this can be seen in the transcript below in Figure 3. At the exact moment Perry achieves the gaze of all four participants (identified by the red arrow in the transcript), she continues her narrative in a more fluid manner, indicating a change in participation framework where she has a unified audience. This shift in participation framework is also evident in Perry's bodily posture. While attempting to secure an audience, she sat tall and moved her pencil in the air in an exaggerated manner (see Figures 2 and 3). After securing each of her peers' gaze, her body relaxes as she rests her chin in her hand (Figure 3, 30 seconds).

**Figure 3.**



The situation of Perry launching her story demonstrates the relevance of the audience's continual presence in the transcription. The non-speaking participants in this scene are neither uniform nor non-agentive. They rather possess differential agency in which they potentially speak, or they write, or they listen to the speaker. This diversification of the audience influences the talk in progress. That Perry is eventually able to achieve the prototypical unified audience depends not only on her performance—and the multiplicity of interactional resources she brings to bear on her performance—but on the audience's agentive choice to participate as a listener.

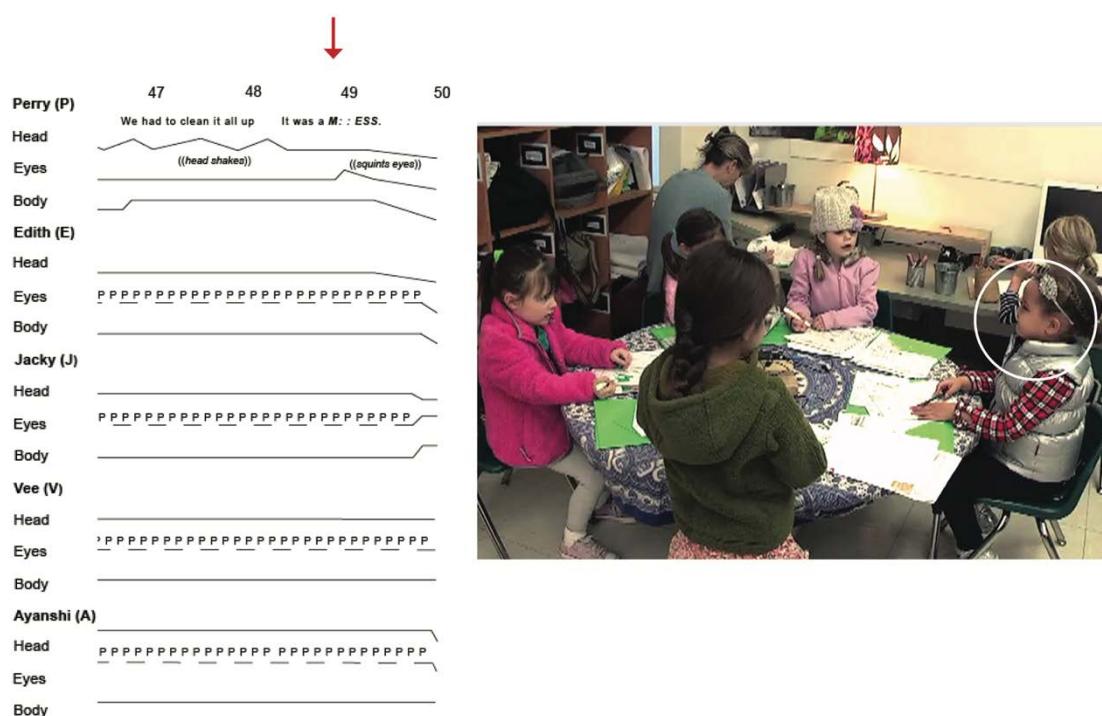
## 6.2 Interactional resources of the audience and story production

As noted by C. Goodwin (1986), "the meaning that the story will be found to have...emerges not from the actions of the speaker alone, but rather as the product of a collaborative process of interaction in which the audience plays a very active

role" (p. 283). Our transcription makes visible this process through its display of the interactional resources the audience brings to bear on the field of action created by the story.

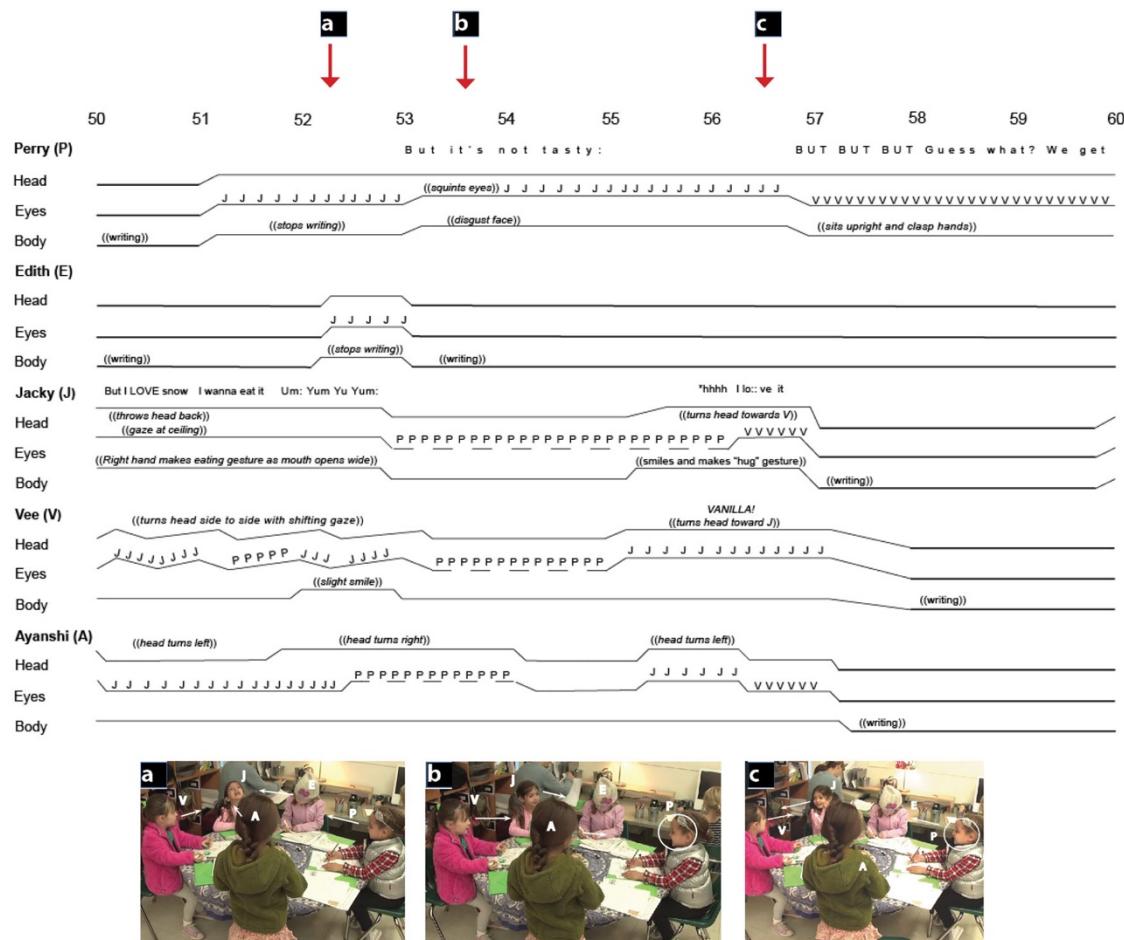
Perry signals the closing of her Santa story—and the participation framework of telling a story to a unified audience—through her statement, spoken with a downward intonation, “We had to clean it all up. It was a mess.” She then returns to her writing. (We refer now to Figure 4 below.) Within this closing, she provides a paralinguistic and nonverbal characterization of how the element of her story where Santa shakes snow off his suit is to be interpreted. Her voice lowers in tone, indicating annoyance, and she gives emphasis to her utterance of the word, “mess,” elongating the “m” so as to create a vibrating sound (49 seconds). She further couples her talk with a head shake and then a squinting of her eyes as she produces the “m” sound (see the image in Figure 4)—actions which serve to add intensity to her negative assessment of the situation (Kendon, 2002). This kinetic and verbal ensemble of meaning effectively conveys that one should interpret the closing of her story as problematic—the mess made by Santa’s entrance was an inconvenience.

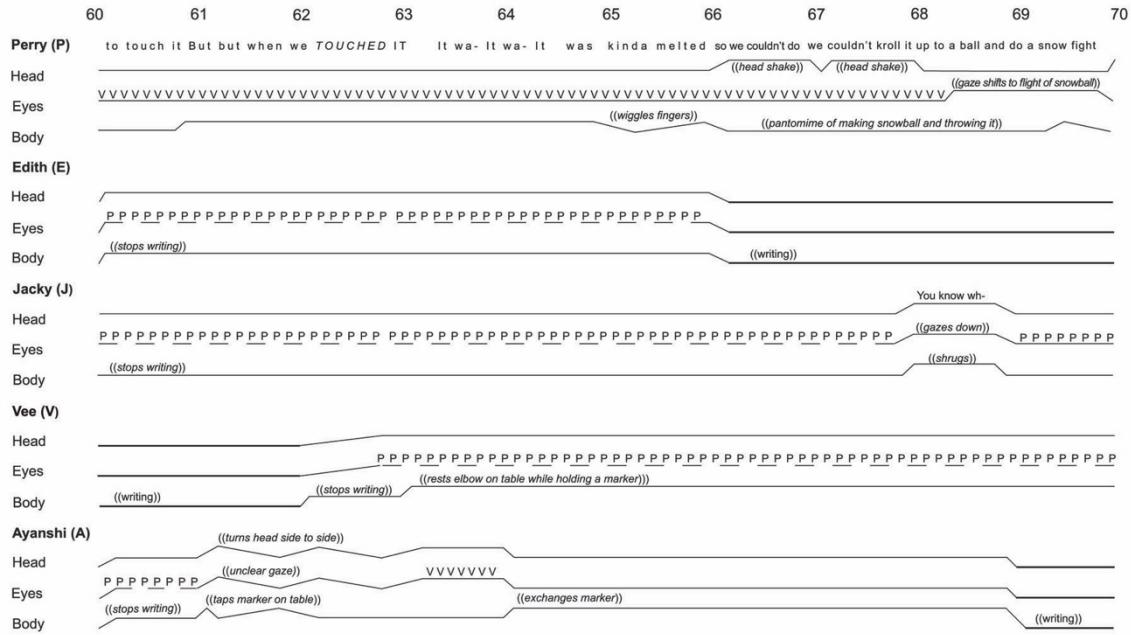
**Figure 4.**



Jacky initiates a shift in footing, however, upon Perry's closing of the story (Figure 5, 50-53 seconds, Image A). Whereas Perry had indicated snow was an annoyance, Jacky uses "but" in the turn initial position to mark a contrastive point of view, namely, that she loves snow (Schiffrin, 1987). Her stance is not only indicated by the lexical and prosodic features of her rapid utterance in which she places emphasis on the word "love" (e.g., But I LOVE snow I wanna eat it), but also through her embodied performance, which displays delight with eating snow. She throws her head back and opens her mouth while mimicking filling her mouth with snow and gobbling it up. As she chomps her teeth, she produces "yum" tokens. In the next turn, Perry, also using the contrastive "but" to indicate an alternative point of view, states: "but it is not tasty"—a position which is elaborated by her performance of a disgust face (Figure 5, 53 seconds, Image B).

**Figure 5.**





At this juncture, Perry and Jacky have established two differing points of view on the palatability of snow. Our transcription shows, however, that beyond these two primary actors, other participants are consequential to this participation framework. Whereas during her storytelling, Perry achieves the sustained gaze of her peers, the girls' gaze now is fragmented. This is most evident with Vee and Ayanshi—both of whom turn their heads side to side from Perry to Jacky as if observing a tennis match. (The reader can see this in the transcript in Figure 5. Notice the rising and falling “gaze line” as Vee’s and Ayanshi’s gazes shift from Perry (“P”) to Jacky (“J”) at 56-57 seconds).

Perry and Jacky’s divergent positions on snow ultimately divide the group into “sub teams.” Our transcription in Figure 5 shows the evolution of this participation framework. Following Perry’s pronouncement that snow is not tasty, Vee’s gaze shifts to Jacky as she exclaims “Vanilla” with a slightly breathy and excited voice (56 seconds). Jacky, who at the same time sharply turns her head to meet Vee’s gaze and utters an exuberant, “I love it,” appears to be able to project her peer’s impending positive assessment (C. Goodwin & M. Goodwin, 1987). Upon achieving a shared gaze, the girls’ concurrent assessments vividly mark their shared affiliation, which positions them in an opposing position to Perry (56.5 seconds, Image C).

Following this assessment sequence, the girls, excepting Perry, return to their writing. Perry opportunistically reclaims the floor as no one is speaking to continue her story. Again, she begins with the story preface, “But but but Guess what?” (57 seconds). The use of the conjunction “but” indicates her story will take a different

trajectory from where she previously left off. She now tells of not being able to form a snowball for a snow fight due to the snow having melted (Figure 5, 63 seconds). As her story appeared to conclude earlier, it is likely she employed this narrative turn as a way to redeem her story after the girls formed an opposing position to her own. In this way, we see the audience's role in constructing a participation framework—through their shared gaze and embodied assessments—that was influential to the speaker's further production of the story (C. Goodwin, 1986).

### 6.3 The “audience” as “speaker”

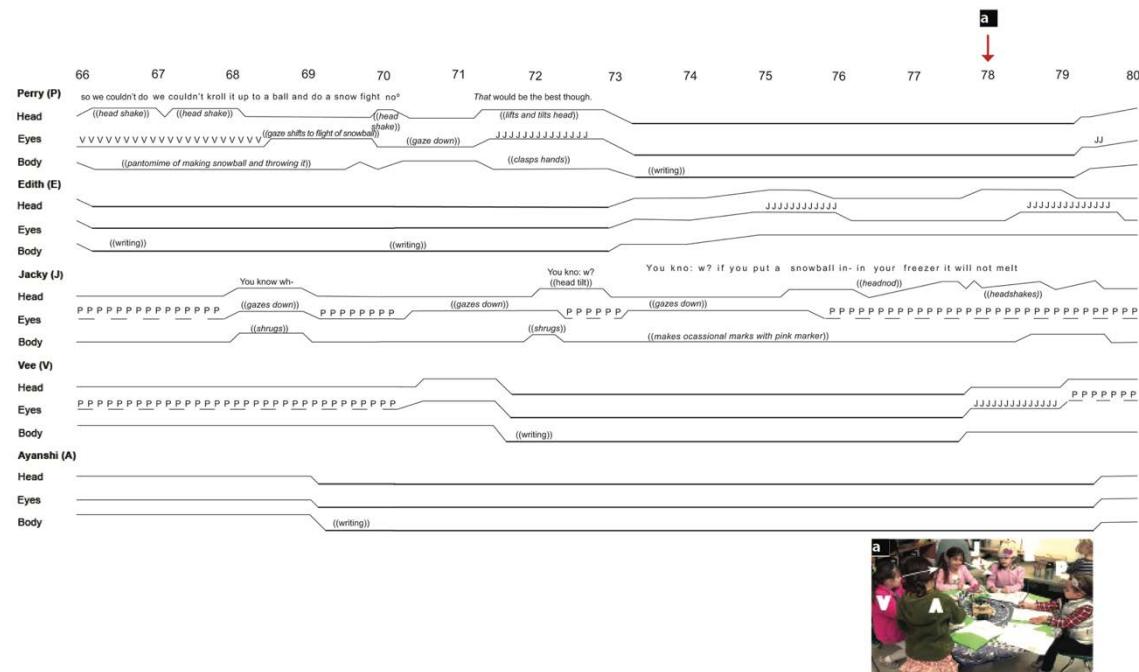
Playscript transcription approaches treat all actors, actions, and utterances as punctual—that is, the transcripts locate these phenomena separately at a specific point in space and time (e.g., the current speaker, utterance, or action). These transcription approaches thus lack the representational power for analyzing the distributed nature of actors and utterances. In this section, we demonstrate how our horizontal transcript makes evident how actors cooperatively construct their action by reusing, indexically incorporating, tying to, and transforming resources provided by others (C. Goodwin, 2018, p. 440). In particular, we are interested in how an audience member is able to mimetically laminate the talk of the speaker in a manner that builds a different action from that originally intended by the speaker.

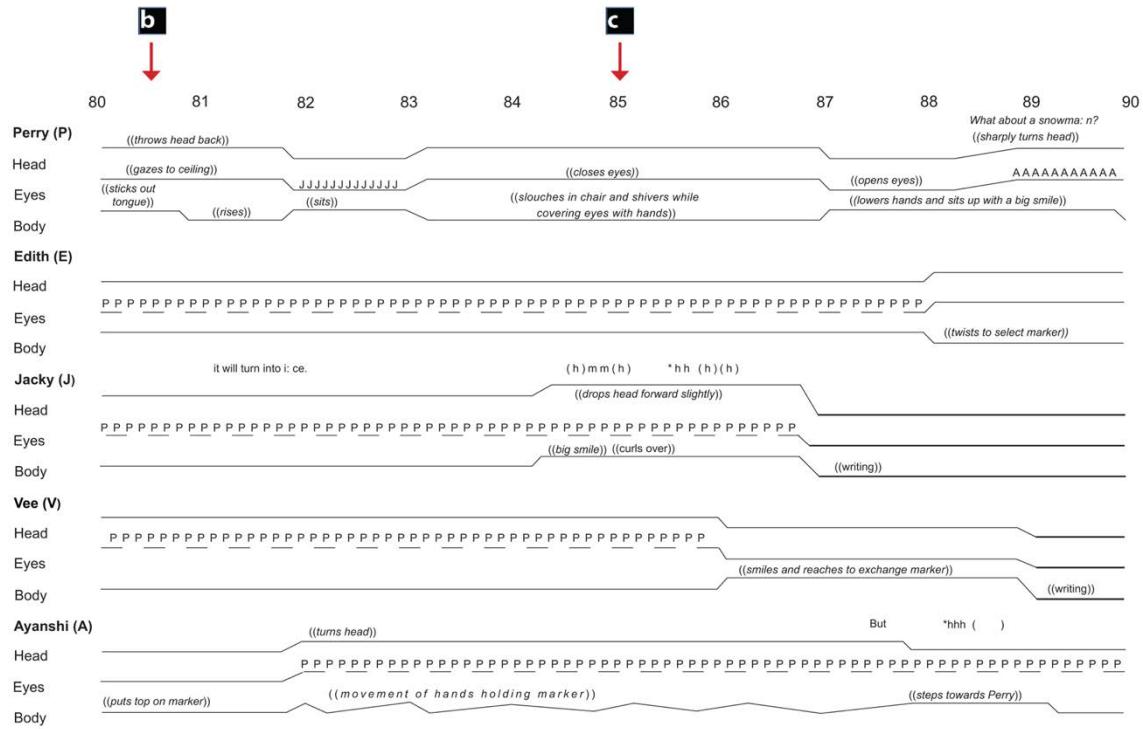
Here we examine the point in the storytelling episode where the traditional narrative morphs into a hypothetical and collaboratively constructed story. This shift of participation framework can be seen in the grammatical structure of the stream of talk as the shift corresponds with the girls' use of the conditional tense. We, however, consider how such a focus on language creates an incomplete understanding of what the girls are accomplishing through their talk *and* interaction.

We continue our examination where we left off in the storytelling scene in the previous section and now refer to the transcript in Figure 6, which can be found below. As Perry closes her story about not being able to make a snowball due to the snow melting, Jacky simultaneously makes a bid for the floor. Jacky repeats the story preface, “You know,” three times, the first two times overlapping Perry’s talk (68-74 seconds). She then continues with a topically relevant statement about melted snow in the conditional tense: “If you put a snowball in- in your freezer it will not melt” (74 seconds). Jacky’s utterance about putting the snowball in a freezer is performed in a matter-of-fact manner. The first clause (if you put a snowball in- in your freezer) is spoken slowly and performed with a serious expression and a head nod and the second clause (it will not melt) also is articulated slowly, with a slight smile and a head shake. Jacky’s statement is followed by a three-second pause; she then continues with a rapidly spoken statement, “it will turn into i:ce” (81 seconds),

followed by a longer, six-second pause during which Jacky chuckles briefly (82-88 seconds).

**Figure 6.**





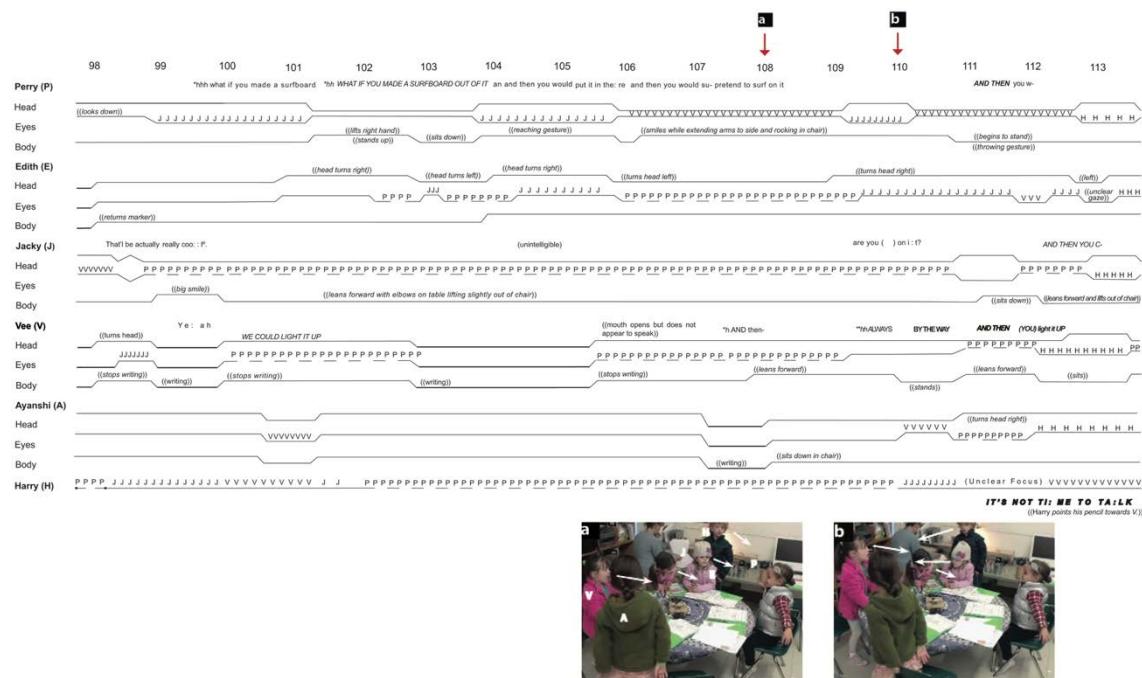
If we transcribed only the talk in this segment, it would appear Jacky “held the floor,” even during her verbal pauses. Observe, however, in our transcript where the participants’ gazes are focused. After Jacky’s initial statement about putting the snowball in the freezer, most participants are gazing at Perry, even after Jacky continues to speak. (This is seen in the transcript beginning at 78 seconds as Edith looks briefly at Jacky and then shifts her gaze to Perry, at whom the other girls are also gazing.) Perry is the primary visual focus of attention because she is using her body to depict the scene Jacky proposes: She throws her head backwards and rises from her chair while sticking out her tongue as if licking an icicle (80.5 seconds, Image A). Her mimesis continues after Jacky’s statement, “it will turn into ice.” Perry slouches in her chair and shivers while covering her eyes with her hands (85 seconds, Image B). Perry’s embodied performance is laminating the talk of Jacky, the speaker, while putting a comedic spin on Jacky’s otherwise matter-of-fact statement. This keying is evident as Jacky responds with laughter (84 seconds). The speaker, in other words, is not an individual but rather divided between the two parties. This phenomenon of “co-operative action”—the way in which action is built

from different parts which are contributed by a single or multiple actors (Goodwin, 2018, p. 441)—aligns with Goffman's notion of a laminated speaker, as well as that of Bakhtin's (1973) and Linell's (2009) conceptions of the dialogic actor. Recognizing how action is distributed and also historically sedimented is essential to understanding how humans as individuals accomplish a range of actions within a larger social matrix. We see this in the present example where Perry's written story about Christmas, which she started writing weeks prior to this scene, acts as the substrate upon which children build an oral story together. This kind of cooperative and distributed action can also be found in professional settings (C. Goodwin, 2018; Hutchins, 1995), children's games (M. Goodwin, 1990), parenting interactions (Andrén, 2017), computer aided communication (Auer & Hörmeyer, 2017) and educational environments (Danish et al., 2020; Johnson, 2017) and is central to creating culturally competent members with specific forms of knowledge and skill within a community of practice (C. Goodwin, 2018).

## 6.4 Embodied turn space

Another example vividly illustrates the importance of capturing co-operative action within transcription. In this case, an actor strategically uses embodied resources to extend the turn space and orchestrate children's gaze.

**Figure 7.**



We refer now to the transcript above in Figure 7. Perry has begun to make a string of propositions as to what one could do with a frozen snowball (which is not included in transcript excerpt). More children, however, are wanting to contribute to the story, and the floor is becoming increasingly competitive. Perry takes a sharp inhale and, gazing at Jacky, says, “what if you make a surfboard” (99.5 seconds). Vee, however, looks up from her writing, gazes at Perry, and with a slightly loud voice proposes “we could light it up,” an utterance which overlaps Perry’s (100 seconds). Perry, now gazing at Vee, restarts her utterance with another sharp inhale. This time she speaks louder until competing talk subsides: “*WHAT IF YOU MADE A SURFBOARD OUT OF IT* an- and then you would put it in the: re and then you would su- pretend to surf on it” (102-108 seconds). Vee continues to make a bid for the floor during Perry’s utterance. She opens her mouth as if to speak, pauses, takes an audible inhale and says, “AND then-” (107 seconds); she pauses again, then inhales sharply and says loudly, “*ALWAYS BY THE WAY*” (109 seconds). These attempts for the floor were executed not only through Vee’s talk but also through her body, adding force to her vocal bid, as she leaned forward and then stood (107-112 seconds).

As Vee upgrades her bid for the floor, it is interesting that she does not gain the gaze of any of her peers other than Perry, who holds the floor. This is the case even after Perry completes her utterance (though Ayanshi briefly gazes at Vee as she stands (110 seconds)). If we examine, however, what Perry’s body is doing *while* this interaction is occurring it becomes clear why this is the case. As Perry tells of making a surfboard and putting it in, presumably, the freezer, she animates her story with her body: She stretches her right arm forward as if placing the surfboard in the freezer and then extends both her arms outward in a surfing pose as she rocks sideways in her chair (100-111 seconds). Notably, she is still surfing after her statement and when Vee raises her voice and stands (109.5 seconds). Perry thus has not relinquished the floor—indeed she has the gaze of all but one of the girls—before she begins her next utterance, “**AND THEN** you w-” which she speaks loudly overpowering Vee’s simultaneous repeated attempt for the floor. Edith, the only child who is not gazing at Perry, is briefly attending to Jacky, who asks Perry if she is on the surfboard (109 seconds).

Perry’s prowess as a narrator is evident in her ability to create imaginative scenarios through her use of language. Her actions, however, are not being built from words alone but, rather, her language is shaped within a larger ecology of mutually elaborating resources. In this instance, arguably, her action of surfing provides greater narrative description (and entertainment) than her statement that she is surfing. She smiles, denoting pleasure, as her undulating body appears to respond to rocky waves. Relatedly, our transcription is able to show this activity in a continuous manner as opposed to it being treated as a footnote, or subsidiary to the utterance, as nonverbal activity commonly is depicted in transcription procedures.

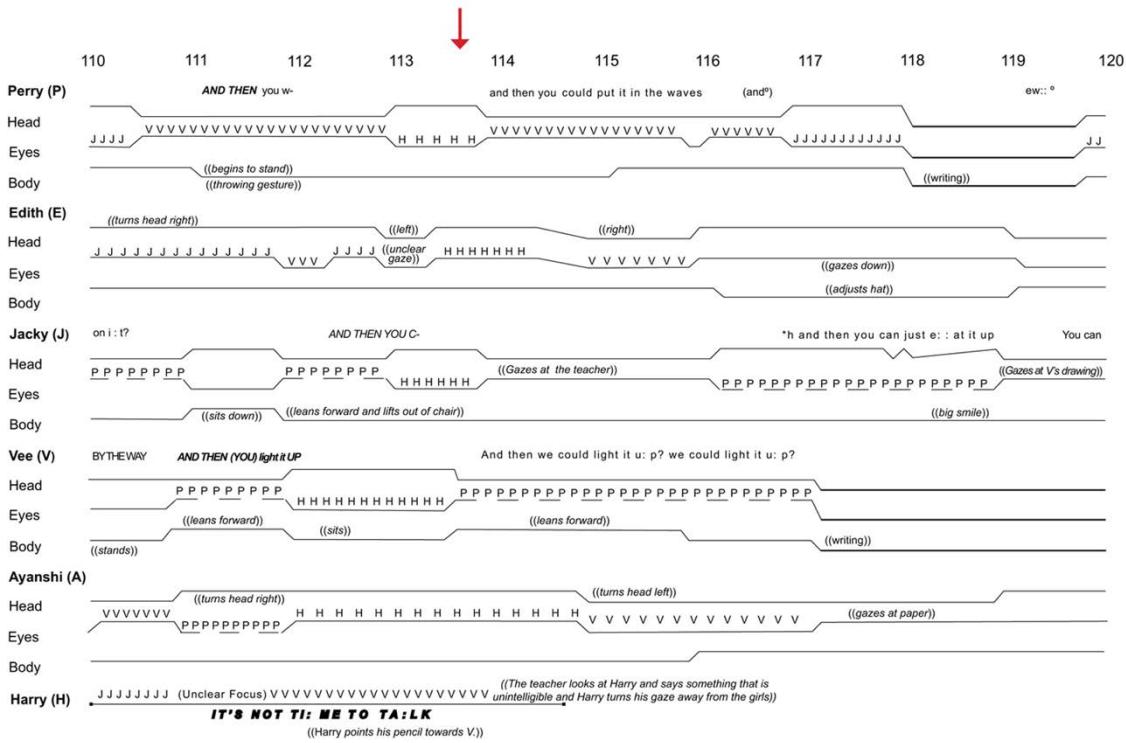
This feature is important, as we see how Perry's action of surfing has captured the attention of her peers and is influencing future action, such as her ability to maintain the floor and control the trajectory of the narrative.

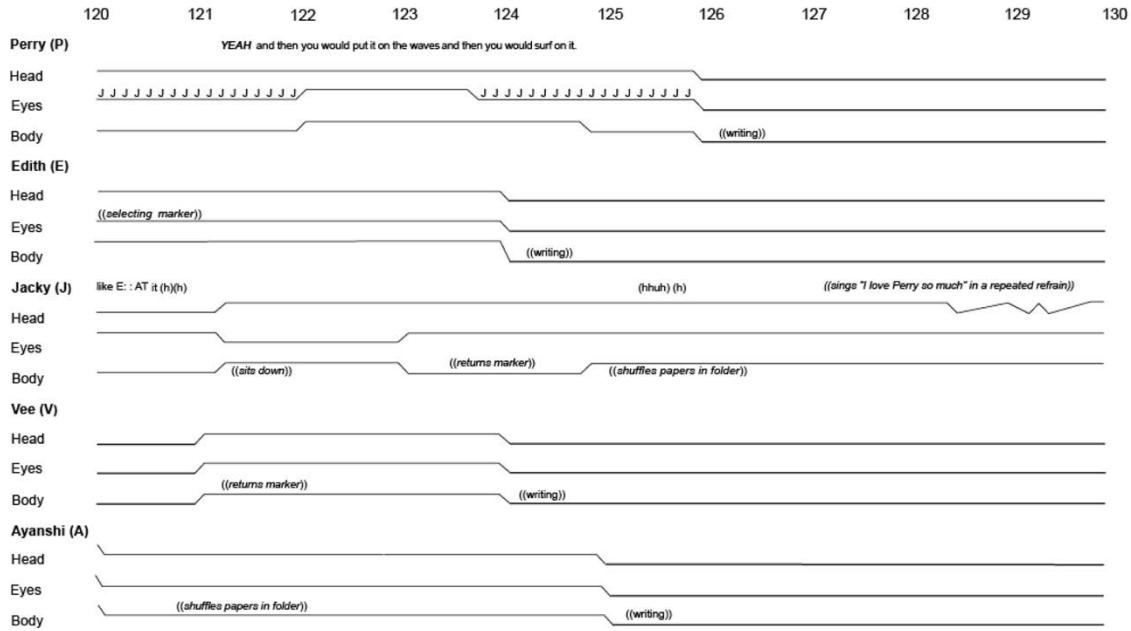
## 6.5 Role differentiation in children's storytelling

The crescendo in action described in the previous section leads to the climax of the children's hypothetical story. Three children simultaneously try to propose their version of events and the young boy, who has been eavesdropping, yells loudly while pointing at Vee, "**It's not ti:me to ta:lk**" (Figure 8, 113.5 seconds). Our transcription is able to display this competitive participation framework of the children's storytelling with all the entailing simultaneous and ongoing activity. In this section, we consider why this analytical representation is important to understanding the lifeworld children construct as part of storytelling, during which they take on various roles within the action (Shultz et al., 1982). To do so, we examine multiple parts of the scene where children tell the hypothetical story as illustrated in the transcripts in Figures 8 9, and 10.

In terms of language use, our transcript makes visible how the children—across three separate speakers—recycle the structure of the previous utterance through their moment-to-moment cooperative action. For example, Jacky's incomplete and loudly spoken utterance "AND THEN YOU C-" (which was interrupted by Harry, who leans in front of her) is recycled (through the repetition of "and then you could") and completed by both Vee and Perry simultaneously, with their respective propositions of lighting the snowball up and putting it on the waves (Figure 7, 112-117 seconds). Jacky, overlapping Perry's talk, then creatively transforms Vee's statement about lighting it up with her utterance: "\*h and then you can just e: : at it up" (116.5 seconds).

**Figure 8.**





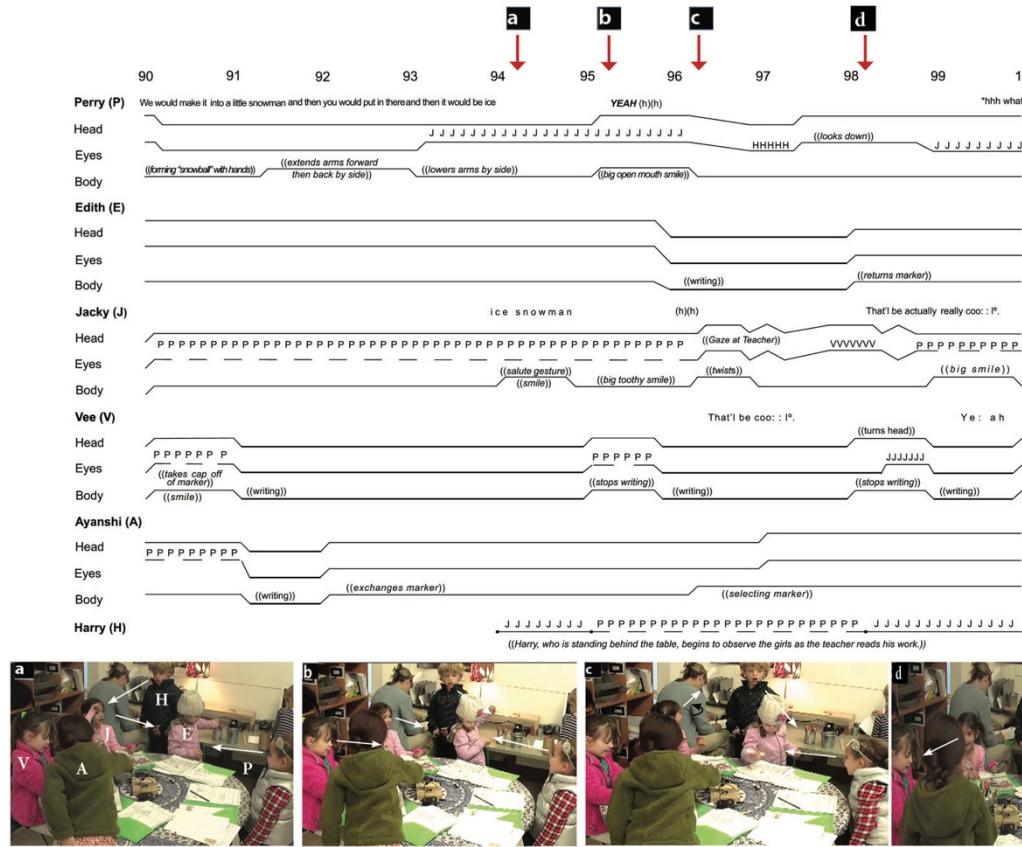
While the structure of the hypothetical narrative (e.g., established storyline and children building from prior talk) invites more democratic participation than a narrative told by a single actor, concomitant with this is increased competition for the floor. This participation framework is perhaps most visible in the fragmented and constantly shifting gaze of participants. Harry, through his uninvited and forceful entry into the conversation, briefly achieves the unified gaze of the girls, as highlighted in the image in Figure 8. Vee, after multiple efforts, is next able to divert attention from Harry and the children briefly attend (as evident in their gaze towards her) to her story proposition about lighting up the snowball ("and then we could light it u:p? light it u:p?") (115 seconds). This attention is remarkable, as her talk is overlapping that of Perry. She does not procure any audible acknowledgement for her contribution, however. Instead, Jacky co-opts Vee's rendering in her own utterance about eating up the snowball (\*h "and then you can just e:at it up...") (116.5 seconds) and Jacky, rather than Vee, receives uptake through the positive assessments from Perry ("ew" and "yeah") (119-122 seconds).

In addition to adding further clauses to the story development, children also participate as speakers in more secondary ways, specifically that of assessing and affirming contributions. Such contributions do not require the speaker to bring together the depth of resources as one would for an extended turn at talk. This, for example, is illustrated by a tripartite assessment among Perry, Jacky, and Vee (which occurs prior to scene where Harry interrupts the girls' story). To consider the children's multiple roles in storytelling we turn to the transcript below in Figure 9. Following Perry's statement about putting the snowman in the freezer, Jacky makes

a salute gesture and, recycling a noun clause from Perry's talk, says "ice snowman," with a smile (94 seconds, Image A). This multimodal package serves as a positive assessment of Perry's story clause. The basis of this interpretation can be found in the multiple assessments which follow. Perry says "yeah" in agreement with Jacky, also providing laugh tokens to key her assessment of the contribution as not only positive, but also funny (95 seconds, Image B); Jacky then marks her shared affiliation by also chuckling (96 seconds). (Laughing is indicated in the transcript with (h) (h) symbols.) Vee chimes in with a softly spoken utterance, "that'll be cool," which Jacky recycles with "that'll be actually really cool" (96-100 seconds). Examining this stream of action in the transcription, we see that for children to contribute these assessments, they did not need to *first* achieve the gaze of all their peers. In Figure 9, the places marked with red arrows, which correspond with images "a," "b," and "c" and "d," show the relationship between turn initiation and gaze. Notice that two children project Perry's turn. Simultaneous with Perry's utterance of "yeah," Harry and Vee gaze at her (95 seconds, Image B).

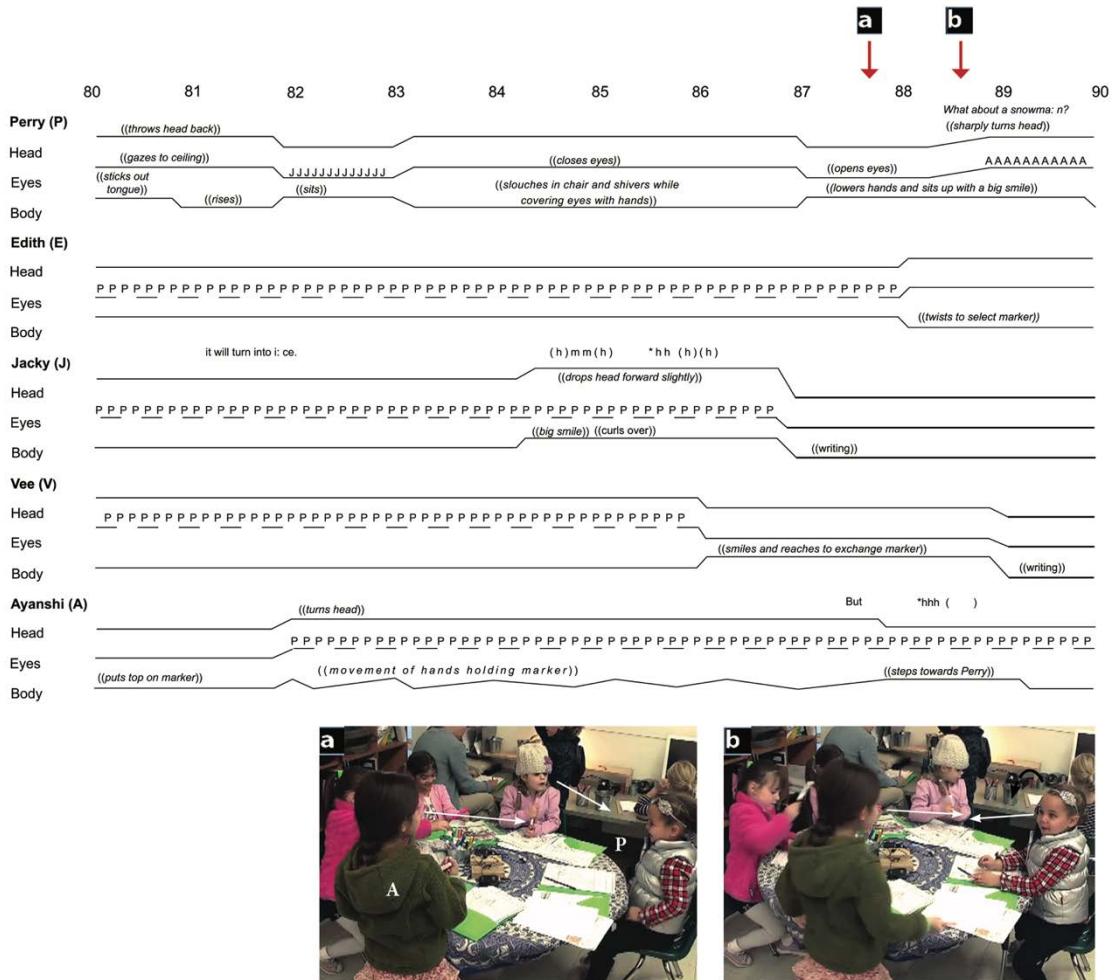
Beyond the resources required for turn initiation, one can observe how such multiparty assessments involve the intricate choreography of children's head movements and gaze as they acknowledge one another's shared perspective. Jacky, for example, seeks the shared gaze of Vee when voicing her assessment and, not finding it, shifts her gaze to Perry, and their eyes meet (97.5 – 100 seconds). (Vee looks at Jacky a moment later at 98 seconds.) The girls are also monitoring both Harry (see "H" in Figure 9), as he begins to attend to their talk, and the teacher, who is reading Harry's paper. Our transcription method is able to track this kind of complex, moment-to-moment change in participation frameworks in a continuous manner rather than treating each aspect of this semiotic ecology as if it occurs in separate dimensions.

**Figure 9.**



What roles do the parties that most exemplify a “listener” play in this storytelling episode? To consider this question, we examine the interaction transcribed in Figure 10 below. Just as Perry closes the sequence where she performs a bodily enactment of freezing ice, Ayanshi gazes at Perry and utters the word “but” (87.5 seconds, Image A). This single lexical contribution suggests Ayanshi has in mind an alternate trajectory for the story. While she identifies the “turn space” for her contribution, she has not accounted for the other participants, whose behavior signals a sequence close as they return to writing. Ayanshi’s gaze towards Perry along with her utterance of “but” (which Perry might interpret as Ayanshi countering the prior version of events she helped narrate) does garner Perry’s attention, however. Perry sharply turns her head to Ayanshi, and overlapping and overpowering her peer’s talk (Ayanshi inhales and appears to continue her clause following “but”), Perry says, “what about a snowma: n” (88.5 seconds).

**Figure 10.**



Ayanshi's chance to be a narrator in this storytelling event is not fully realized. By documenting her continuous participation in the scene, however, we accomplish, for one, an accounting of the varied ways the listener influences the speaker (as discussed throughout this manuscript) along with a more analytically accurate version of how multiparty, cultural activity is conducted. The young children's roles within their peer community are different and multiple, with some children, who are less confident storytellers, observing the activity and potentially gradually "pitching in" more over time (Rogoff, 2014). This is observed in this instance with Ayanshi's ongoing listening activity as she writes and her interpolation of the word "but" into the emerging story. Our transcript thus shows the cultural storytelling activity with the social ecology of its enactment, where children participate in a manner that is more primary, secondary, tertiary, and so on, and in which these varied activities of the participants relate in ways that take account of one another (Erickson, 2010).

## 7. Conclusion

This paper has argued for further development of horizontal approaches to video-based transcribing of the moment-by-moment conduct and organization of social interaction face to face in ways that make salient the simultaneity of conjoint social action as well as its sequentiality, so as to avoid the limitations of “linguocentrism” in playscript transcription formats. What we have presented here is one attempt at a new approach. We do not intend to advocate for this particular example as a new “standard” for horizontal transcription. Indeed, because new transcription and new theory develop together and the field of video-based studies of interaction is still in the early stages of an “embodied/ecological” turn, it is too soon to adopt some new “standard” even if some might think that a standard approach was desirable.

The example presented here has implications for our understanding of semiosis—how humans manage, through their varied forms of participation, to do meaning in conjoint social action. What Perry and the other children are able to do as speakers is deeply implicated in what their interlocutors are doing as listeners. As the “whiles” of interactional activity is emphasized in transcription and analysis, it becomes even less credible than it has already been that *language* communicates meaning as an autonomous system of contrasting features, as Saussure initially claimed and others have claimed since. The word never stands by itself. Rather, the word, together with all other semiotic media, communicates situated meaning from moment to moment within an interactional ecosystem. And within that ecosystem of continuous mutual influence, it can justly be said that the word is made flesh.

As well, there are also implications for social theory in this example of transcription and its presumptions about the conduct of interaction as fundamentally ecological. Among these implications are ones concerning the nature of human agency. If interaction is a matter of enacted ecology, then the agency of participants in it is not simply a matter of individual choice, nor does the boundary of an individual’s actions stop at the level of that person’s skin. Rather, human agency is distributed within the local situation of conjoint activity (Enfield & Kockelman, 2017). Perry and the other children cannot simply do whatever they want as they engage with one another—each participant must continually take action within the environment constituted by the actions of the other participants. As with other aspects of post-modern ontology, this de-centers the individual social actor. And as Charles Goodwin (2018) has observed in his magisterial account of the *co-operative* nature of interaction, we “*inhabit*” one another’s actions: “as we inhabit each other’s actions we move through lived time together, while co-operatively transforming what is occurring there” (p. 477).

In sum, we have tried to treat the listener with full regard in our transcription in a way that is consonant with Goodwin’s understanding of co-operation in social interaction.

The transcription and discussion show how children collaboratively construct an imaginative and evocatively enacted set of oral narratives. This is one attempt at horizontal transcription across a constant timeline. It is presented in the hope that others' attempts might follow.

## Acknowledgements

Sarah Jean is deeply appreciative of Mike Rose for his encouragement and support to her throughout the long transcribing and writing process for this manuscript. A few weeks prior to his passing in August of 2021, he offered extensive commentary on a completed draft of the paper, which shaped and refined our analysis and arguments. We also are thankful for the helpful critique provided by two anonymous reviewers.

## References

Alkemeyer, T., Brümmer, K., Pille, T. (2017). Intercorporeality at the motor block: On the importance of practical sense for social cooperation and coordination. In C. Meyer, J. Streeck, & J. Scott Jordan (Eds.), *Intercorporeality: Emerging socialities in interaction* (pp. 203-236). UK: Oxford University Press.

Andrén, M. (2017). Children's expressive handling of objects in a shared world. In C. Meyer, J. Streeck, & J. Scott Jordan (Eds.), *Intercorporeality: Emerging socialities in interaction* (pp. 105-141). UK: Oxford University Press.

Arbeau, T. (1589/1967). *Orchésographie*. Initially published in Langres. Translated and republished in New York: Dover Press.

Auer, P., & Hörmeyer, I. (2017). Achieving intersubjectivity in augmented and alternative communication (AAC): Intercorporeal, embodied and disembodied practices. In C. Meyer, J. Streeck, & J. Scott Jordan (Eds.), *Intercorporeality: Emerging socialities in interaction* (pp. 323-360). UK: Oxford University Press.

Bakhtin, M. (1973). *Problems of Dostoevsky's poetics*. Ann Arbor, MI: Ardis.

Bucholtz, M. (2000). The politics of transcription. *Journal of Pragmatics*, 32, 1439-1465.

Calkins, L. M. (2001). *The art of teaching reading*. New York: Longman.

Churchill, E. (2022). The education of attention in enskillment. *Journal of Pragmatics*, 193, 155-172.

Condon, W. and Ogston (1967). A segmentation of behavior. *Journal of Psychiatric Research*, 5, 221-235.

de Souza, D. K., Park, S. H., Wei, W., Zhan, K., Bolden, G. B., Hepburn, A., Mandelbaum, J., Mikesell, L., & Potter, J. (2021). The Gratitude Opportunity Space: The timing of gratitude expressions in object passes. *Social Interaction. Video-Based Studies of Human Sociality*, 4(1).

Duranti, A. (2006). Transcripts, like shadows on a wall. *Mind, Culture, and Activity*, 13(4), 301-310.

Ehlich, K. and Rehbein, J. (1976). Halbinterpretive Arbeitstranskriptionen. *Linguistische Berichte*, 45, 21-41.

Ehlich, K. (2014). HIAT: A transcription system for discourse data. In J. Edwards & M. Lampert (Eds.), *Talking data: Transcription and coding in discourse research*, (2<sup>nd</sup> ed., pp. 123-148). New York: Psychology Press.

Enfield, N. and Kockelman, P. (2017). *Distributed agency*. Oxford UK: Oxford University Press.

Erickson, F. (1982). Money tree, lasagna bush, salt and pepper: Social construction of topical cohesion in a conversation among Italian-Americans. In D. Tannen (Ed.), *Analyzing discourse: Text and talk* (pp. 43-70). Washington, D.C.: Georgetown University Press.

Erickson, F. (1986). Listening and speaking. In D. Tannen and J.E. Alatis, (Eds.), *Language and linguistics: The interdependence of theory, data, and application* (pp. 294-319). Washington, DC: Georgetown University Press.

Erickson, F. (2009). Musicality in talk and listening: A key element in classroom discourse as an environment for learning. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality: Exploring the basis of human companionship* (pp. 449-463). Oxford, UK: Oxford University Press.

Erickson, F. (2015). Oral discourse as a semiotic ecology. In D. Tannen, H. Hamilton, and D. Schiffrin (Eds.), *The Handbook of Discourse Analysis* (2<sup>nd</sup> ed., pp. 422-446). Hoboken, NJ: John Wiley & Sons.

Erickson, F., & Shultz, J. (1982). *The Counselor as Gatekeeper: Social Interaction in Interviews*. New York: Academic Press.

Goffman, E. (1964). The neglected situation. In Gumperz, J. and Hymes, D. (Eds.) *The ethnography of communication*. Special issue of *American Anthropologist*, 66(6) Part 2, 133-136.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.

Goodwin, C. (1986). Audience diversity, participation and interpretation. *Text-Interdisciplinary Journal for the Study of Discourse*, 6(3), 283-316.

Goodwin, C. (2018). *Co-operative action*. Cambridge, UK: Cambridge University Press.

Goodwin, C., & Goodwin, M. H. (1987). Concurrent operations on talk: Notes on the interactive organization of assessments. *IPrA papers in pragmatics*, 1(1), 1-54.

Goodwin, M. H. (2021). The interactive construction of a hug sequence. In A. Cekaite and L. Mondada (Eds.), *Touch in social interaction* (pp. 27-53). New York: Routledge.

Goodwin, M. H. (1990). *He-said-she-said: Talk as social organization among black children* (Vol. 618). Bloomington, IN: Indiana University Press.

Goodwin, M. H., & Cekaite, A. (2018). *Embodied family choreography: Practices of control, care, and mundane creativity*. New York: Routledge.

Goodwin, M. H., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, 62, 51-75.

Gumperz, J. J. (1982). *Discourse strategies* (No. 1). Cambridge University Press.

Heath, C., & Christian, H. (1986). *Body movement and speech in medical interaction*. Cambridge University Press.

Heath, C., & Luff, P. (2021). Embodied Action, Projection, and Institutional Action: The Exchange of Tools and Implements During Surgical Procedures. *Discourse Processes*, 58(3), 233-250.

Hutchins, E. (1995). *Cognition in the Wild* (No. 1995). Cambridge, MA: MIT Press.

İkizoğlu, D. (2019). "What did it say?": Mobile phone translation app as participant and object in family discourse. *Journal of Pragmatics*, 147, 1-16.

İkizoğlu, D. (2021). *Agency and participation in multilingual family interaction*. [Doctoral dissertation, Georgetown University]. ProQuest dissertation publishing.

Iwasaki, S., Bartlett, M., Manns, H., & Willoughby, L. (2019). The challenges of multimodality and multi-sensoriality: Methodological issues in analyzing tactile signed interaction. *Journal of Pragmatics*, 143, 215–227.

Johnson, S. J. (2015). *The social and cognitive worlds of young children reading together*. [Doctoral dissertation, University of California, Los Angeles]. ProQuest dissertation publishing.

Johnson, S. J. (2017). Multimodality and footing in peer correction in reading picture books. *Linguistics and Education*, 41, 20-34.

Kamunen, A. (2019). How to disengage: suspension, body torque, and repair. *Research on Language and Social Interaction*, 52(4), 406-426.

Kendon, A. (1970). Movement coordination in social interaction: Some examples described. *Acta Psychologica*, 32, 101-125.

Kendon, A. (1990). Some context for context analysis. In A. Kendon (Ed.), *Conducting interaction: Patterns of Behavior In Focused Encounters* (pp. 15-48). New York: Cambridge University Press.

Kendon, A. (2002). Some uses of the head shake. *Gesture*, 2(2), 147-182.

Klein, G. (2020). *Pina Bausch's dance theater: Company, artistic practices and reception* (Vol. 56). Aachen, Germany: Transcript.

Kress, G., & Van Leeuwen, T. (2001). *Multimodal discourse. The modes and media of contemporary communication*. London: Arnold.

Labov, W. (1972). *Language in the inner city: Studies in the Black English vernacular* (No. 3). Philadelphia, PA: University of Pennsylvania Press.

Leeds-Hurwitz, W. (1987). The social history of the natural history of an interview: A multidisciplinary investigation of social communication. *Research on Language and Social Interaction* 20(1-4), 1-51.

Lerner, G.H., Zimmerman, D.H. and Kidwell, M. (2011). Formal structures of practical tasks: a resource for action in the social life of very young children. In E. Streeck, C. Goodwin and C. LeBaron (Eds). *Embodied interaction*:

*Language and body in the material world* (pp. 44-58). Cambridge: Cambridge University Press.

Laurier, E. (2014). The graphic transcript: Poaching comic book grammar for inscribing the visual, spatial and temporal aspects of action. *Geography Compass*, 8(4), 235-248.

Linell, P. (2009). *Rethinking language, mind, and world dialogically: Interactional and contextual theories of human sense-making*. Charlotte, NC: Information Age.

McDermott, R. (1976). *Kids make sense: An ethnographic account of the interactional management of success and failure in one first grade classroom*. [Unpublished doctoral dissertation, Stanford University].

Merlino, S. (2021). Professional touch in speech and language therapy for the treatment of post-stroke aphasia. In A. Cekaite and L. Mondada (Eds.), *Touch in Social Interaction* (pp. 197-223). New York: Routledge.

Meyer, C., & Wedelstaedt, U. V. (2020). Achieving Competitive Hardness: The Preference for Hard Touch as Practical Accomplishment in Professional Handball. *Social Interaction. Video-Based Studies of Human Sociality*, 3(1).

Mode transcription bank (nd). University College London. Retrieved January 23, 2022, from <https://mode.ioe.ac.uk/category/transcription-bank/>

Mondad, L. (nd). *Conventions for transcribing multimodality*. Retrieved January 23, 2022, from <https://www.lorenzamondada.net/multimodal-transcription>

Mondada, L. (2019). Transcribing silent actions: A multimodal approach of sequence organization. *Social Interaction. Video-based Studies of Human Sociality*, 2(1).

Mondada, L., Monteiro, D., & Tekin, B. S. (2021). The tactility and visibility of kissing: Intercorporeal configurations of kissing bodies in family photography sessions. In A. Cekaite and L. Mondada (Eds.), *Touch in social interaction* (pp. 54-80). New York: Routledge.

Norris, S. 2004. *Analyzing multimodal interaction*. London: Routledge Falmer.

Ochs, E. (1979). Transcription as theory. In E. Ochs and B. Schieffelin. (Eds.), *Developmental pragmatics* (pp. 43-72). New York: Academic Press.

Ochs, E., & Capps, L. (2009). *Living narrative: Creating lives in everyday storytelling*. Cambridge: Harvard University Press.

Poylanyi, L. (1982). Linguistic and social constraints on storytelling. *Journal of Pragmatics*, 6, 509-524.

Plowman, L., & Stephen, C. (2008). The big picture? Video and the representation of interaction. *British Educational Research Journal*, 34(4), 541-565.

Routarinne, S., Heinonen, P., Karvonen, U., Tainio, L., & Ahlholm, M. (2020). Touch in achieving a pedagogically relevant focus in classrooms. *Social interaction: Video-based Studies of Human Sociality*, 3(1).

Rogoff, B. (2014). Learning by observing and pitching in to family and community endeavors: An orientation. *Human development*, 57(2-3), 69-81.

Sacks, H. (1967). Transcribed lectures. March 9th, University of California, Irvine.

Sacks, H., Schegloff, M., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, 50, 696-735.

Scollon, R. (1982). The rhythmic integration of ordinary talk. In D. Tannen (Ed.), *Analyzing discourse: Text and talk* (pp. 335-49). Washington, D.C.: Georgetown University Press.

Scollon, R. (2001). Action and text: towards an integrated understanding of the place of text in social (inter) action, mediated discourse analysis and the problem of social action, In R. Wodak & M. Meyer (Eds.), *Methods of critical discourse analysis* (pp. 139-183) London, UK: Sage Publications.

Scheflen, A. (1973). *Communicational structure: Analysis of a psychotherapy transaction*. Bloomington, IN: University of Indiana Press.

Schiffrin, D. (1987). *Discourse markers* (No. 5). Cambridge, UK: Cambridge University Press.

Sicoli, M. A. (2020). Modality, Multimodality. In J. Stanlaw (Ed.) *The International Encyclopedia of Linguistic Anthropology* (pp, 1-5). Hoboken, NJ: John Wiley and Sons.

Slembrouck, S. (2007). Transcription—the extended directions of data histories: a response to M. Bucholtz's 'Variation in Transcription'. *Discourse Studies*, 9(6), 822-827.

Streeck, J. (1984). Embodied contexts, transcontextuals, and the timing of speech acts. *Journal of Pragmatics*, 8(1), 113-137.

Shultz, J. J., Florio, S., & Erickson, F. (1982). Where's the floor: Aspects of the cultural organization of social relationships in communication at home and in school. P. Gilmore and A.A. Glathorn (Eds.), *Ethnography and education*, 88-123. Washington, D.C.: Center for Applied Linguistics.

Tomlinson, K. (1735). *The art of dancing, explained by reading and figures*. London: Printed for the author.

Tufte, E. (1990). *Envisioning information*. Cheshire CT: Graphics Press.

von Laban, Rudolf (1928). *Schriftanz: Methodik, orthographie, erlaeuterungen (German)*. Vienna: Universal Edition.

Weber, M. (1922/1978). *Economy and society: An outline of interpretive sociology*. G. Roth & C. Wittich (Trans.). Berkeley CA: University of California Press.