



Visualizing Emergent Turn Construction: Seeing Writing While Speaking

Greer, Tim

Nanbu Zachary, Matthew K T

(Citation)

The Modern Language Journal, 106(S1):69-88

(Issue Date)

2022-03-03

(Resource Type)

journal article

(Version)

Version of Record

(Rights)

© 2022 The Authors. The Modern Language Journal published by Wiley Periodicals LLC on behalf of National Federation of Modern Language Teachers Associations, Inc.
Creative Commons Attribution-NonCommercial License

(URL)

<https://hdl.handle.net/20.500.14094/0100483092>



Visualizing Emergent Turn Construction: Seeing Writing While Speaking

TIM GREER¹  AND ZACHARY NANBU² 

¹Kobe University, School of Languages and Communication, Graduate School of Intercultural Studies, 1-2-1 Tsurukabuto, Nada-ku, Kobe, 657-8501, Japan

²Kobe University, Graduate School of Intercultural Studies, 1-2-1 Tsurukabuto, Nada-ku, Kobe, 657-8501, Japan

This study draws on multimodal conversation analysis to emically account for moments in second language (L2) English interaction in which speakers appear to be visualizing text as they talk. One way they do this is by slotting out elements of a turn-in-progress in the air, shifting their hand in a slotting gesture from left to right as they say each word to display to their recipient that they are visualizing certain elements of the turn. In other cases, participants use their fingers to ‘write’ elements of the turn-in-progress on their palms or in the air. The embodied practices of visualizing a turn component by component as it is formulated therefore make public the temporality of its in situ grammatical production. These multimodally accomplished visualizations also provide the speaker with access to a recalled text that helps them produce the spoken equivalent. The study suggests that English-as-a-foreign language (EFL) learners may therefore support their spoken interaction by visualizing written grammar or lexical items, and that multimodal practices such as the precision-timed deployment of gaze and gesture make a seemingly intrapsychological process like visualization a social matter. The data are taken from a corpus of 94 video-recorded paired discussion tests among EFL learners whose first language (L1) was Japanese.

Keywords: second language interaction; interactional competence; oral assessment; turn construction; multimodal conversation analysis

LANGUAGE AND ITS DEVELOPMENT ARE both deeply rooted in sociality and the way people interpret their worlds. The structures of language emerge through social interaction and become routinized within a repertoire of practices for accomplishing social actions (Hall, 2018; Pekarek Doehler & Balaman, 2021; Pekarek Doehler & Eskildsen, 2022, this issue). In spoken interaction, the way a turn unfurls across time (its temporality)

can be observed via its step-by-step development and the way its design is adapted in situ according to local contingencies. The syntactic emergence of a turn projects and constricts various next items and sequentially due follow-up turns, and so accounting for linguistic structures in relation to timing and temporality is a central concern for interactional linguistic research (Mushin & Pekarek Doehler, 2021).

At the same time, social action is also achieved through co-occurrent embodiment, including gaze, gesture, and proximity, and these embodied features both shape and become accountable to the specific contingencies of the local interactional ecology (Mondada, 2021). In addition, emergent body movements can enable recipients to project the development of a turn event before it has ended (Depperman & Günthner, 2015; Keevallik, 2018; Mondada, 2021). Embodied elements of a learner’s interactional repertoire are often drawn on to help accomplish meaning, particularly in the early stages of learning a second

The Modern Language Journal, 106, S1, (2022)

DOI: 10.1111/modl.12748

0026-7902/22/69–88 \$1.50/0

© 2022 The Authors. *The Modern Language Journal* published by Wiley Periodicals LLC on behalf of National Federation of Modern Language Teachers Associations, Inc.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.



language (L2). On occasion, for example, speakers will produce a series of beat gestures (McNeill, 1992) or parsing gestures (Kendon, 2004; Streeck, 2008) during the realization of a turn-constructional unit (TCU). Unlike iconic gestures that are designed to convey semantic information, these gestures instead assist in illustrating speech structure, pacing its delivery and drawing emphasis via movements of the hand timed more or less simultaneously with words or phrases. Among the beginning learners of English in our data, we also see hybrid versions of these gestures that more clearly depict written forms of a turn-in-progress. At times these gestures seem to represent syntactic slots or entire words, since the speaker positions their hand with the thumb and forefinger apart and moves it progressively across the table as they produce their turn. As they do this, their gaze is usually averted away from their partner and/or toward the gesturing hand. The embodied progression stops during any silence, suggesting that the speaker is “doing visualizing”¹ a written version of the sentence as it unfolds. In other cases, speakers use their fingers to ‘write’ elements of the turn-in-progress on their hands or in the air by tracing the shapes of recalled orthographic forms. These multimodal visualization practices both signal to the recipient that a search is underway and provide the speaker with step-by-step access to the temporal emergence of a publicly imagined text that helps them arrive at the spoken equivalent. Our study suggests that English-as-a-foreign-language (EFL) learners can therefore support their spoken interaction by visualizing written grammar and lexical items. Participants orient to the temporal organization of grammar in interaction via multimodal practices such as the precision-timed deployment of gaze and gesture, and these resources make unfolding action trajectories projectable.

We begin with an overview of previous research on visualization within interaction and, after outlining the study’s dataset, we offer a sequential analysis of seven cases from our collection to illustrate the microgenesis of turns-at-talk as they emerge in and through interaction. The overarching aim of the study, then, is to account for how novice English users deploy multimodal resources to temporally render grammar a shared object of visualization.

VISUALIZATION IN INTERACTION

Our focus in this study is not on text visualization in a literal sense, such as that experienced by so-called tickertape synesthetes (Holm,

Eilertsen, & Price, 2015) or when readers decode visibly available writing on a page in front of them (Gough, 1972). In fact, we adopt an analytically agnostic stance toward what the participant is actually seeing during moments of text visualization. Our interest in this phenomenon is instead purely interactional, or “outside the skull” (Kasper, 2009). That is, our focus is on the grammar–body interface (Mushin & Pekarek Doehler, 2021) and how speakers can, through the carefully timed deployment of embodied resources, display that they are seeing (or imagining or recalling or searching for) lexical or syntactic elements of a developing turn. These temporally grounded practices have real-time consequences for the interaction, displaying that the speaker is working on their formulation of the turn and thus observably accounting for breaks in progressivity. We are therefore not so much concerned with what the individual is actually seeing, but the way they “do seeing” and how that becomes locally established as a relevant practice for social interaction, particularly for speakers using an unfamiliar language.

Like cognition or identity, vision is something that undeniably takes place within the head of an individual. Even so, it is also regularly made public via the vocal or aural modality (i.e., the organization of talk) and the visuospatial modality of embodiment (Stivers & Sidnell, 2005). Speakers visibly orient to certain things as relevant, and people around them properly demonstrate they understand this to be the case. In this way, vision can be considered a socially distributed phenomenon. Interactants can make use of semiotic resources like spoken utterances, embodiment, and the physical environment, incorporating them into laminated actions (Goodwin, 2013), gesture–speech ensembles or utterance packages (Kendon, 2004), or complex multimodal gestalts (Mondada, 2014) to invite their recipients to see and understand an action as it unfolds. Interaction is therefore made up of both auditory and visible events (Goodwin, 1994): In co-present situations, interaction can best be understood by considering visibly available elements like space, gaze, and bodily movements that mutually constitute it (Duranti, 1992). For example, part of what makes up a person’s “professional vision” (Goodwin, 1994) is not just the ability to see things, but to make those things ‘seeable’ to others via the organization of social action: An archeologist on a dig can trace, inscribe, and highlight certain sections of the ground to make them noticeable to an apprentice (Goodwin, 1994), or a violin teacher can use her thumb and



forefinger to show her student which part of the bow to use (Nishizaka, 2006). Such environmentally coupled gestures thus become part of a laminated action that seamlessly comprises talk, space, bodily movement, gaze, and the like—and this may ultimately become the medium through which learning takes place.

Visualizing text is therefore one particular type of seeing. By design, text is an artifact of the bodily movements that produced it, whether via the affordance of a pen, brush, or printer (Mondada & Svinhufvud, 2016). The path that a line of ink inscribes on paper constitutes a historical remnant of earlier human action, yet the purpose of any inscribed object is partly a matter of how it is put to use (Day & Mortensen, 2017). When writing happens through the movement of a finger in the air, the product is highly ephemeral, leaving the viewer to interpret the strokes in conjunction with any other available semiotic resources in real time. Such air-writing is fairly common among Japanese speakers (Thomas, 2015), whose complex logographic script (*kanji*) does not possess a simple means of spelling out words verbally. Instead, speakers often trace out orthography in the air or on the hand to facilitate repair during conversation (Cibulka, 2013) or to reflect on the way a kanji character should best be written (Arano, 2020). Understanding such writing is, therefore, primarily a temporally grounded spatiovisual act that straddles the border between speaking and writing. Our study takes up Cirillo's (2019) call for further scrutiny of the relationship between text and talk and responds to Stevanovic & Monzoni's (2016) appeal for a move away from logocentric categorizations of social action (i.e., analysis that focuses only on language). In doing so, it offers a systematic account for the complex interplay of semiotic resources that come together when novice English users represent text through gestures as they talk.

PURPOSE OF THE STUDY

By undertaking a systematic temporal analysis of interaction between novice English learners in a paired discussion test, we aim to emically account for instances of the embodied practice of text visualization. Our interest in the target phenomena emerged from extensive “unmotivated looking” (Sacks, 1992) at video recordings of the test data, which led us to pose two questions. First, what are the spoken and embodied practices that inform a recipient that a speaker is visualizing text, and what repercussions does this have for the ongoing talk? Second, what can text visualization

in interaction tell us about the real-time manner in which novice EFL language users plan and produce a turn-at-talk?

In exploring these questions, our analysis will contribute to the growing body of research within interactional linguistics that uncovers the interplay of language structure, temporality, emergence, and projection (Auer, 2015; Deppermann & Günthner, 2015; Hopper, 2011) and respond to Mushin & Pekarek Doehler's (2021) call for further investigation into the grammar–body interface. Our microgenetic analysis of emergent turns will suggest that spoken interaction can be buttressed by the visualization of written grammar, and that multimodal practices, such as the precision-timed deployment of gaze and gesture display, render such textual visualization a publicly available and thus interactionally salient social practice.

DATA

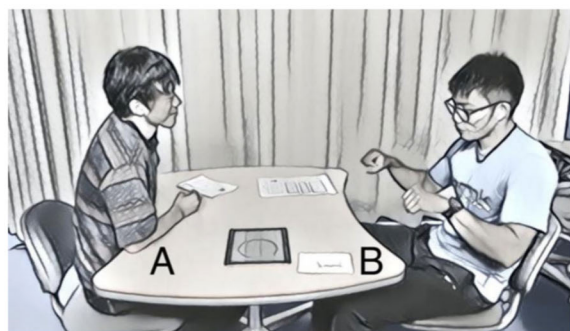
The interactional data to be examined all come from a corpus we call the Kobe Test of Oral Proficiency (KTOP; Greer, 2019). This collection of 93 peer-to-peer discussion tests was video recorded among first- and second-year students at a university in western Japan as one of the assessment requirements for a core English communication class. All participants gave written consent for their recordings to be studied and the research satisfied ethical approval conditions from the national research body, the Japan Society for the Promotion of Science (JSPS). Even though most Japanese university students have learned English at secondary school for 6 years, many of them do not actively speak it, so we refer to these participants as novice language users. Although the average proficiency level of the participants in our broader dataset is B1 on the Common European Frame of Reference for Languages (CEFR) scale, those who feature prominently in our analysis are more likely at the A2 level.

At the point the recordings were collected, the students had participated in 7 weeks of an oral English class that focused on developing their spoken fluency through pair and group discussions, including topics such as travel, marriage, share-housing, and extended family. Pairs of test-takers were randomly assigned one of these topics by selecting a card just prior to the start of the test. They were then required to talk freely in English about that topic for 4 minutes. The test took place in a room near their regular classroom and the class teacher was not present: A proctor video recorded the interaction and the teacher



FIGURE 1

The Seating Arrangement During the Kobe Test of Oral Proficiency [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]



later graded it according to a performance-based rubric that included fluency, accuracy, and complexity. The test accounted for 20% of the students' final grade.

The test takers were seated across from each other, as shown in Figure 1. Throughout the analysis, and across different dyads, we will identify the test taker on the left as "A" and the person on the right as "B." On the table, there was a prompt card that consisted of only one or two words—the topic that the students randomly selected (e.g., "travel" or "jobs").

METHOD

Conversation Analysis

Following Firth & Wagner's (1997) call for greater sensitivity to participant-centered perspectives within the study of second language acquisition (SLA) research, conversation analysis (CA) has been widely used to investigate L2 conversation and interactional competence in the classroom and beyond. CA constitutes both a method of analysis and a sociological theory in its own right (Heritage, 1984). Its aim is to uncover the interactional practices used by participants to organize talk-in-interaction and account for these practices in terms of the participants' own orientations. To this end, CA's analytic concern is with the microdetails of talk. Recurring interactional practices are gathered via audio and video recordings and sorted into collections of cases based on their sequential and action-relevant properties. CA findings involve a detailed, emic description of the focal phenomena. For the current analysis, an initial collection of candidate cases was identified from the dataset; these were transcribed in detail, and common features of these excerpts were identified. The collection included 12 cases of visualizing individual words through air-writing

and 14 cases in which the speaker mapped out syntactic units. Of these, seven representative excerpts will be analyzed here.

Transcription

Spoken elements of the interaction have been transcribed according to Jeffersonian conventions (Jefferson, 2004) with embodied features indicated in gray. Following a simplified version of Mondada's (2018) approach, each embodied tier is identified with the participant's initial and a code indicating the locus of embodiment (e.g., "-gz" for gaze, "-px" for proximity, "-bh" for both hands, and so on). The onset of the embodied action is located relative to the talk tier via a horizontal bar (|). See the Appendix for further details.

ANALYSIS

Our analysis will focus on representative excerpts from our collection to account for three visualization practices: (a) flagging a word, (b) air-writing in forward-oriented repair, and (c) visualizing syntax through slotted gestures. Since space does not permit a detailed analysis of our complete dataset, we have selected sequences that most clearly demonstrate the focal phenomena. We begin by examining a case in which writing text in the air seems to flag the word for particular attention within the turn. Next, we account for similar visualization practices within forward-oriented repair (i.e., word-search) sequences, suggesting that these smaller writing gestures design the turn as a solitary word search. Finally, we consider text visualizations during longer syntactic constructions by showing how speakers can visibly slot out a sentence word by word across an environmentally available space. Common to each of these practices is that the speaker makes their visualization of text available to the recipient in order to demonstrate they are dealing with some difficulty in the timely production of the turn-in-progress, and the grammar-in-interaction is laminated through temporally sensitive embodiment.

Flagging a Word

One interactional locus in which air-writing can be found in our data is the flagging of a word for particular emphasis. This may be, for example, because the speaker is designing that segment of the turn as a somewhat atypical usage or as potentially unknown to the recipient, or to mark a Japanese word as possibly violating the institutionally mandated language medium (English).



EXCERPT 1

Marimo

- 01 B do you have any pets?
 02 (0.6)
 03 A |ah: (.) |no I haven't.
 a-rh |waves left
- 04 A |↑ah- | (.) | demo (.)
 but
 a-gz |to B |down-----
 a-rh |raises to chin
 b-hd |nods-----|
- 05 A |eh: i:- in fi:ve yea::(.)rs |ago,
 a-rh |thumb points right
- 06 B |uh[un]
 b-hd |nods
- 07 A |[I:] I have.u (.) |mari(h)mo(h)?
 a moss ball
 a-rh |raises, extends index |air-writes まりも (marimo)
 a-px |leans in slightly |sit back
 fig |2A |2B
- 08 B |↑↑marimo?
 b-hd |nods
- 09 A |ma(h)[ri mo]
 a-hd |nods
 a-rh |air-writes まり (mari)
 fig |2C
- 10 B |[ma(h)]rimo .heh hah ha
 b-hd |nods



Like “air quotes” (Cirillo, 2019), this type of air-writing can pragmatically serve to highlight the vagueness of a referent or distance the speaker from its truth value. Cibulka (2013) referred to episodes of air-writing as framing and suggested that they differ from air quotes in that they are recognizable as text and therefore “underline the specificity of the given item” (p. 176). Excerpt 1 is a case in point: Speaker A writes the Japanese word *marimo* ‘moss ball’ in the air as he says it.

In line 1, B asks A if he has any pets, and after initially responding in the negative (3), A

then goes on to qualify this by saying he used to own a *marimo*, a kind of moss ball found in lakes throughout Japan and sometimes sold in stores as a low-maintenance “pet.”² However, A displays his orientation to the atypicality of his answer in line 7 via a brief delay and then laughed-through delivery of the word “marimo” with upwardly intoned “try-marking” (Sacks & Schegloff, 1979).

It is at this point that A also writes “marimo” in the air. During the first part of his turn in line 7, he raises his hand to roughly chest height with his index finger extended where it is clearly



visible to B, readying his hand to air-write. As he does this, he leans forward slightly and says “I-I have,” extending the vowel and following it with a micropause at a grammatically incomplete point. Both his grammar-in-interaction and his embodied practices therefore make projectable to the recipient that the upcoming word deserves particular attention. At the end of line 7, A completes the turn with “marimo,” quickly flicking his index finger in the air as he says it. He writes vertically in what appears to be Japanese, depicting it from his own perspective rather than from B’s. His precision-timed delivery of the audible and visual elements make clear which word he is writing in the air, while the act of writing it flags his use of the Japanese word as marked within the institutional context of the English discussion test. Although clearly iconic, this gesture also shares some important characteristics with beat gestures, in that it seems to be emphasizing “something the speaker feels [is] important with respect to the larger discourse” (McNeill, 1992, p. 40).

Taken together, speaker A’s laminated delivery of “marimo” in line 7 works to index the potential unexpected or laughable nature of his answer. In the next turn (8), B does indeed treat “marimo” as unexpected, with her high-pitched other-repetition initiating repair in a news-marked manner, occasioning another round of repetitions (9, 10). A’s confirmation in line 9 is accompanied by a partial reprise of his air-writing gesture in which he only completes the first two characters of the word. This second retrospective iteration further couples his embodied action to the word and ties this repetition to what came before. Because the gesture is not reproduced in full and its strokes are more subtle, its recognizability as writing is contingent on anaphoric reference to A’s earlier, more pronounced version. It is not that A needs to see it or negotiate it for himself but more that he is treating it as part of the joke (i.e., the depiction of a plant as a pet), and indeed it does receive more open laughter from B in line 10.

This form of air-writing is clearly designed for the recipient, produced in a space and manner that is visible to both participants. The speaker launches a turn segment (“I have...,” 7) that grammatically projects a slot in which the turn-completing object becomes sequentially due. When the (L1 Japanese + air-writing) formulation hearably completes that turn in progress, the recipient is invited to see the air-writing as that of the spoken element with which it co-occurs. In short, this sort of air-writing is formulated as part of a multimodal gestalt that flags part of the turn as potentially problematic for the recipient.

Visualizing Text in Forward-Oriented Repair

In other cases, the act of writing with a finger or planning out a sentence on the table can be seen as primarily facilitating the turn construction for the current speaker, with the recipient orienting to it in that way by refraining from interrupting or through minimal uptake tokens upon its completion. The interactional locus in which this is most often found in our dataset was in word-search sequences, or what we refer to more broadly as forward-oriented repair (Schegloff, 1979), since such moments may also involve a search for content (Hasegawa, 2017) and the like. As outlined by Schegloff, Jefferson, & Sacks (1977), the repair organization is a set of interactional practices for dealing with trouble in talk. While such trouble can be due to speaking, hearing, or understanding, the repair is most often backward-oriented, in that it seeks to retrospectively rectify some repairable that an interactant has identified in previous talk, even immediately prior elements of their own turn. In contrast, forward-oriented repair sequences target a trouble source in yet-to-be-produced talk, that is, the timely production of a word or some prospective linguistic element. As such, forward-oriented repair is usually instigated by the current speaker (self-initiated) and therefore the repair proper is often completed by that same speaker, since they hold the floor. Although a common feature of L1 talk (Goodwin & Goodwin, 1986; Hayashi, 2003), such word-search sequences have also been shown to provide opportunities for language learning among novice language users (Brouwer, 2003; Duran, Kurhila, & Sert, 2019; Koshik & Seo, 2012), where the situated, collaborative practice of searching for a word can lead to “learnables” (Eskildsen, 2018) that novice language users may later enlist in their talk.


Unsurprisingly, gaze direction and embodiment have been found to be highly relevant components of forward-oriented repair. When a speaker looks away from their audience after initiating forward-oriented repair, they are signaling that they are attempting a solitary word search, and if they return their gaze to the recipient without displaying that they have found the word, the recipient can then take this as an invitation to join the search by proffering a candidate solution (Dressel, 2020; Goodwin & Goodwin, 1986). Recipient collaboration can also be mobilized through the use of prepositioned gestures, which provide a projection space within the turn (Hayashi, 2003). The hands can be used to indicate an intended recipient or produce a



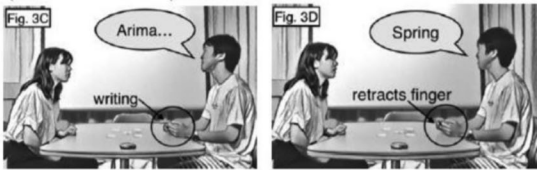
EXCERPT 2

Arima Spring

01 B |uh- kita ku |is famous |for uh:
 North ward
b-gz |up-----|~~to A-->
fig |3A |3B



02 |arima, |arima[::]|spring.u?
b-rh |index swish |index up and down slightly
b-gz |-----|~~up-----|~~upward right
b-hd | |tilts left
a-hd |nods
fig |3C |3D



03 A |[ah-]
a-hd |nods 3 times

04 A arima spring.u,

gestural representation of the searched-for word (Taleghani-Nikazm, 2015). We view air-writing as a particular sort of depictive gesture that focuses on the orthographic nature of the word rather than its physical likeness.

In Excerpt 2, B enacts an episode of air-writing as he searches for the word “spring,” which is part of his in situ translation of a Japanese place name (Arima Spring).

B’s telling is briefly put on hold as he searches for the word “spring” as part of a place name. His gaze during most of line 1 is up and away, but he turns back briefly to A as he produces the hesitation marker “uh.” In line 2, he accesses the Japanese component of the name “Arima” with an initial swish of his index finger, but then repeats that word with a vowel extension as he again directs his gaze up and away from the recipient in a “thinking face” (Goodwin & Goodwin, 1986, p. 57) while air-writing with his index finger. The combined (or laminated) effect of these spoken and visual turn components shows the recipient that B is doing thinking at this moment, and in particular that he is visualizing the written form of the place reference. Although B’s TCU is hearably incomplete, the recipient displays recognition of “Arima” by producing a change-of-state token in partial overlap in line 3.

While B’s word search is a solitary one, this does not mean that A is not participating. Because of B’s consistent shifts in gaze leading up to his word search, A was able to project and display alignment by maintaining her gaze on him without interrupting, thus allowing his search to continue. As Goodwin & Goodwin (1986) noted, when a recipient gazes at a speaker engaged in a word search, they evidence the fact that such practices are accomplished systematically by both parties via complex displays of co-participation in the ongoing talk. B eventually arrives at the word “spring,” but tilts his head left as he utters the word with an upward intonation, indexing uncertainty with his word choice (Seo & Koshik, 2010). However, in the next turn, A treats this as recognizable by repeating “Arima Spring,” and the talk moves forward. Just as air-writing became a means of marking the word “marimo” in Excerpt 1, B’s air-writing here projects issues with the word “spring.” His gestures are partially self-directed, helping him visualize the word while also making the trouble available to his interlocutor. Compared to *marimo* (Excerpt 1), the air-writing in this excerpt is much lower and smaller, reflecting the largely solitary nature of the word search.

Although the air-writing in these two examples was associated with Japanese words or names, this



FIGURE 2

The Slotting Gesture Is Held Out Horizontally Over the Surface of a Table



is by no means the case in our broader collection. We found it used also in relation to English words in other conversations (e.g., “so,” “marriage,” “university”) or to write numerals when they were calculating, translating, or remembering someone’s age. In Excerpt 2, B is most likely translating “spring” from the original place name (*onsen*), and so the air-writing could be helping him to arrive at the English, or at least inform his interlocutor that he is working on it.

Visualizing Syntax Through Slotted Gestures

Whereas the word searcher in the previous excerpt was clearly wiggling an extended index finger to do writing—and thereby orienting to a single word—in other situations, participants use a slotting gesture and the affordance of the table to plot segments of sentences word by word. Canonically, this is done with thumb and forefinger held apart as shown in Figure 2 although other variations are possible.

Rather than the shapes of letters, in these cases each hand movement represents an entire word, and the speaker seems to be gesturally mapping grammar by parsing the sentence into its components as they produce it. An initial case of this phenomenon can be seen in Excerpt 3. Here, A and B are discussing jobs, and A slots out a sentence across the table as he formulates it.

In this excerpt, we see the speaker processing language at the sentence level. He does not use his index finger as a symbolic pencil to write each word, but instead enlists the whole hand to ‘place’ words onto the table, thus orienting to their syntactic relationship as he positions them in an ordered manner. In line 1, he begins a turn segment

(part of a telling sequence) that he links to earlier talk with “but I.” He stops mid-sentence and shifts his gaze away from the recipient, then upward as he utters a seemingly self-addressed repair initiator (“hm?”), which projects trouble within the turn-in-progress (see Steinbach & Thorne, 2011, on the public face of self-directed speech). It is at this point that he mobilizes his hand to begin setting out the sentence both grammatically and spatiotemporally. In line 2, he holds his right hand in a slotting gesture as he (re)produces the turn-initial element “I,” placing it at the far-left side of the table. In other words, the recipient (B) is invited to see what A places on the table as the word “I.” While still looking away, A then goes on to repeat the gesture word by word as he moves his hand across the table, shifting his gaze from the left side of the table to the timer, which is placed on the right side of the table. Produced with prosodic breaks between each component of the sentence, the slotted gesture suggests that the speaker’s gaze is orienting to the textual nature of the turn-in-progress. He is doing seeing each word in his emergent syntactic construction and, moreover, is asking the recipient to see him doing it that way. This practice therefore gives rise to the microgenetic development of not only emergent syntactic structure but also of a shared object of imagination.

Another example of this phenomenon can be seen in Excerpt 4, in which the same speaker (A) is explaining his father’s job. The text visualization comes in line 11.

As in the previous excerpt, in line 11 we again observe Speaker A slotting out a sentence across the desk. After telling B that his father is now making GPS systems for cars (1–9, omitted), A then goes on to say that his father used to make televisions (10–11). This then-and-now juxtaposition is accomplished through A’s use of similar grammatical and lexical choices, as well as a repetition of some key gestural components. In line 10, A makes a swiping gesture with his right hand as he utters several hesitation markers and shifts his gaze up and away from B. This swipe may be indicating that the previous talk (about his father’s current work) has finished and that he is moving on to a different topic (his father’s former job), or it may be foregrounding the phrase that comes immediately after it, “one years ago” (11), in that it seems to be deictically placing it in the past. If this is the case, A’s gesture may be the sort that helps him access the word he is searching for.

In line 11, A again taps out the sentence “one years ago he make a television” on the desk. As shown in the framegrabs, his gaze is up and away

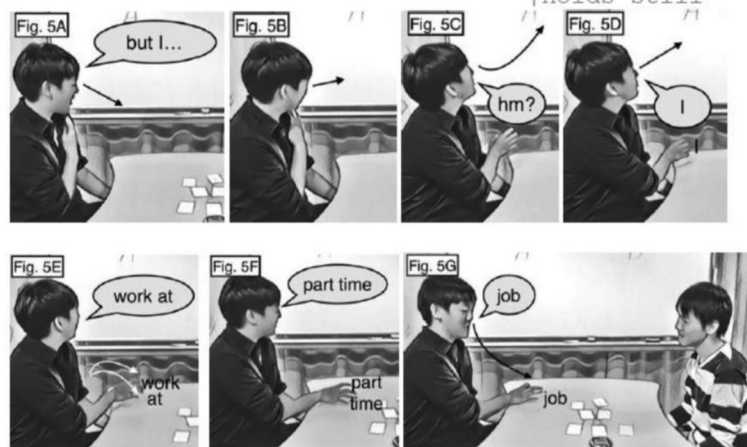


EXCERPT 3

Part-Time Job

01 A |but |I: |(1.0) |work- |hm? (0.8)
 fig |5A |5B |5C
 a-gz |down~|left-----|~up/left
 a-rh |touches left shoulder
 |rh to table
 |lifts rh slightly

02 A → |I- |(0.8) |work |at |part time |job
 fig |5D |5E |5F |5G
 a-gz |up/left-----|~table/timer-->
 a-rh |touches table far left
 |rh motionless
 |touches table left
 |touches table near left
 |touches table center
 |holds still



from the recipient during the early part of his turn production, again suggesting that he is visualizing the sentence as he says each word. His hand taps the desk on “years” and “ago,” but the left-to-right progression is less pronounced in this instance, since he is touching his sleeve during “one” and “make.” However, when he comes to the key content word of the sentence, the turn-final “television,” he first glances down and draws something on the desk before turning to B and saying “television” as he makes a circle on the desk. What he actually draws during the 0.5-second silence is unclear, but there are two possibilities: He could be writing the word “TV” as the speaker did with “marimo” in Excerpt 1, or he could be sketching a quick icon of a television to help him access the word. Either way, it is worth noting that A’s gesture changes subtly at this point, going from the slotting gesture (that represents the placement of a full word) to a single index finger that is used to spell out a word to be highlighted. His circling motion in the same part of the desk as he produces “television” supports this hypothesis that he is treating that as worthy of particular note.

Continuing on directly after this, Excerpt 5 also shows how A treats the table as a space for visualizing a textual version of his spoken turn-in-progress.

In this excerpt, A uses an open-handed version of the slotted gesture to depict key words of his talk. In lines 14 and 15, he tells B where his father works, again looking away for the majority of the turn as he hesitantly produces the initial grammar-intensive sections of the sentence and then turning to his recipient as he produces the turn-final token, “Panasonic” (15). As he does so, he again taps the desk with his thumb and forefinger roughly 10 centimeters apart and in the same vicinity of the desk where he had been visualizing text just earlier. There is nothing in particular about this gesture that would suggest an iconic representation of the word “Panasonic,” so the recipient is left to understand the gesture as representing the word itself, or the slot in the visibly mapped sentence it occupies.

A then goes on in lines 17–19 to reprise his earlier sentence, adding that his father makes



EXCERPT 4

Television

10 A |eh:t | (2.1) |°eh:to° (0.9)
 HM HM
 fig | 6A | 6B
 a-gz |~~down|~~up/left ----->
 a-rh |swipes|~~~~~|swipes



11 → |one |years |ago |he make a | (0.5) |television.
 fig | 6C | 6D | 6E | 6F | 6G | 6H
 a-gz |-----|~~down|~~B----->
 a-rh |beat| beat |beat x2 |sleeve |draws |circles



12 | (0.4)
 a-gz |----->

13 B |oh-hohn
 fig | 6I
 b-hd |nods



televisions and computer displays (i.e., monitors). He looks down at the desk during the first part of his turn in line 17, then again looks to the left as he taps visualized words onto the table. The framegrabs in line 19 show how he spaces each element of the second half of his turn (“television and eh computer display”) across the table from left to right. His gaze meets B’s on “television,” the word that ended A’s earlier version of this turn, but then he looks away as he produces the incremental “and” and then taps out “computer display” on the desk as he redirects his gaze

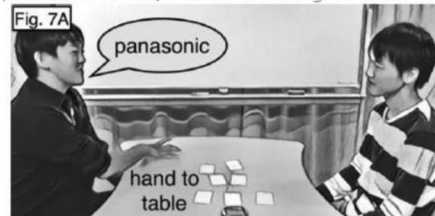
to the recipient. The temporal development of the turn in progress is mediated not only through the grammar-in-interaction, but also through the speaker’s visibly available embodied attention to each turn element as he accesses it. Those moments when he is looking away or at the table seem to be most directly related to the text visualization in terms of individual cognitive processing, but the overall effect of gesturally setting out the sentence on the table is to show the recipient that he is dealing with the difficulty of accurately producing the turn.



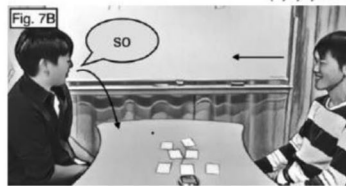
EXCERPT 5

Panasonic

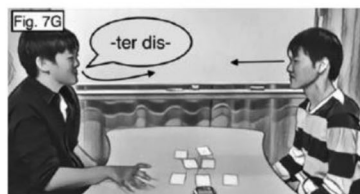
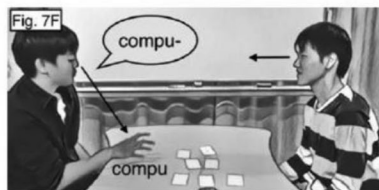
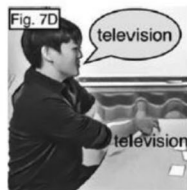
- 14 A |oh et, |he: (0.7) °|tch° work |at:
HM
a-gz |>>~~down/right----|~~up-----|down-->
a-rh |-----|,,,under table----->
a-lh |adjusts RH sleeve
- 15 A → |(0.2) |>panasonic.<
fig |7A
a-gz |-----|~~B-----
a-rh |-----|~~to table, index finger and thumb apart



- 16 B |oh
b-hd |neck forward, nods
- 17 A → |so: |[|he:] |(0.5) make (0.8) |e:h (0.6)
fig |7B |7C
a-gz |down|~~left-----|~~up----->
a-rh |LH arm |~~to table, index finger and thumb apart
|,,,to lap-----|~~to table->



- 18 B [nice(job)]
- 19 A → |>television< |and |(0.9) eh::|compu|ter |dis|play.
fig |7D |7E |7F |7G |7H
a-gz |-----|down-----|~~B-----
a-rh |slot left |slot |holds |slot slot |right



- 20 | (0.5)
a-gz |~~down
- 21 B |°ohn°
a-gz |~~timer
a-rh |,,,to lap-->
- 21 A |°un°
a-gz |~~timer-->
b-hd |nods



Despite this difficulty with formulation, Speaker A has been highly effective at holding the majority of the floor at this point, and his embodied practices have played a major role in this achievement. Having receipted his telling, this might be one point at which B could take a longer turn to reciprocate, but A again self-selects to continue his telling and text visualization practices.

A begins his turn in line 23 in Excerpt 6 with the connective “but” but immediately self-initiates open-class repair with *un* (“huh?”), which delays the turn progressivity as the activity again shifts to a word search. Interestingly, he places his hand on the left side of the table at this point, as if he was about to tap out another sentence across it, but during the 0.7-second silence that follows, his hand does not move, again providing evidence to suggest that this gesture represents the turn-in-progress. When he resumes the sentence with “I see,” his hand moves too. On “I,” he does a swishing gesture that seems to be initiating backward-oriented repair. Retrospectively, it is possible to interpret this as an insertion repair: The “but” is put on hold while A inserts “I see his work,” and then “but” is repeated (25) to continue the turn. In line 24, A again holds his hand over the table as his vocal and embodied conduct show B that he is dealing with some trouble in formulating the turn. The immobilization of the gestural progression signals that A has not yet visualized what he is trying to say, and the co-occurrent hesitation markers provide verbal evidence for this. Conversely, in line 25, it appears that he has arrived at a candidate completion to the turn and therefore continues to tap on the desk for “his work” before holding it again during the 0.9-second gap of silence.

In line 27, A’s turn reaches completion with “what he doing,” but he immediately self-repairs this to “what he was doing.” As can be seen in the framegrab, the initial completion is produced while establishing mutual gaze, but as A moves to repair it, he again looks away and sweeps his right hand to the left of the table, apparently as if erasing the words he has just said. This sort of backward-oriented repair, therefore, also provides evidence to suggest he is visualizing the words as he says them. However, it is unlikely that he is actually seeing words in the air; our claim is that he is making a display of visualizing the turn elements as he produces them, showing the recipient that he is working on getting the sentence out. The hand held in slotting position can project more to come or, with a move toward the left of the table, retrospectively address earlier parts of his

turn and thus becomes a locally established practice for the temporal organization of his talk.

Backward-Oriented Repair

These sorts of backward-oriented repair sequences are frequently combined with forward-oriented repair in our dataset as the interactants work to formulate a turn in English, and as we have seen, their visualization of turn elements can become perceptible along with retrospective grammatical insertions, replacements, and abandonments. In Excerpt 7, speaker A is specifying the name of a place and abandons his turn to formulate it with simpler grammar, synchronizing it with a combination of hand gestures that become an integral part of the multimodal gestalt (Mondada, 2014).

This segment extends the talk we analyzed in Excerpt 3 (1) as A goes on to specify the name of the place he works (3–4). He begins to say, “it takes so...” (3), but then abandons this turn and replaces it with “this is a(t) Kumon” (4). His hand movement in line 3 is hesitant, wavering back and forth from left to right as he produces the initial turn segment. It then stops midair as he utters the cutoff “so” and directs his gaze away from B, indicating some sort of interactional trouble. When he begins speaking again in line 4, he also starts tapping out each word across the table, initially with his gaze averted but subsequently redirected toward B by the end of the turn. His hand stops on “a” and he shakes it slightly during the 0.5-second gap, again suggesting hesitancy. Finally, the turn ends with the key word “Kumon” (the name of a school). When he produces this final token, A’s hand is no longer open-palmed as it was while he was tapping out each word; instead, he positions it as though he is gripping a pencil and gives a couple of quick swishes above the desk as he says it. Although not absolutely clear from the video recording, one way to interpret these strokes is as writing, in particular the Japanese katakana character for “ku” (ク), the first symbol in the word “Kumon.” Similar to the way the speaker in Excerpt 1 flagged a word by writing it, A’s intonation and gesture here seem to be marking “Kumon” as possibly unknown to B. If this is indeed the case, this instance represents a combination of each of the visualization practices we have discussed: tapping across the desk during forward-oriented repair, pausing and rewinding the progression of such tapping during moments of backward-oriented repair, and flagging a word with a written gesture.



EXCERPT 6

His Work

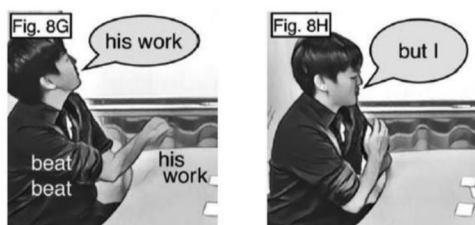
23 A → |but |un? |(0.7) |I see |(0.3) |tch
 huh?
 fig |8A |8B |8C |8D
 a-gz |----|~~left|-----
 a-rh |~~~|slot |LH sleeve |swish |sleeve |slot to table



24 → |(1.1) |h- |un?
 huh?
 fig |8E |8F
 a-gz |-----|~~up/left-->
 a-rh |-----|slot |holds above table-->



25 → |his |work, |(0.9) |but I
 fig |8G |8H
 a-gz |-----|~~down/center-->
 a-rh |slot |beat |hold---|~~to L shoulder-->



26 | (0.3) |don't know
 fig |8I
 a-gz |----->
 a-rh |----->



27 → |what |he doing. |>what he< was |doing.
 fig |8J |8K
 a-gz |----|~~B-----|~~up/left----|~~B-
 a-rh |-----|swipes across table R to L



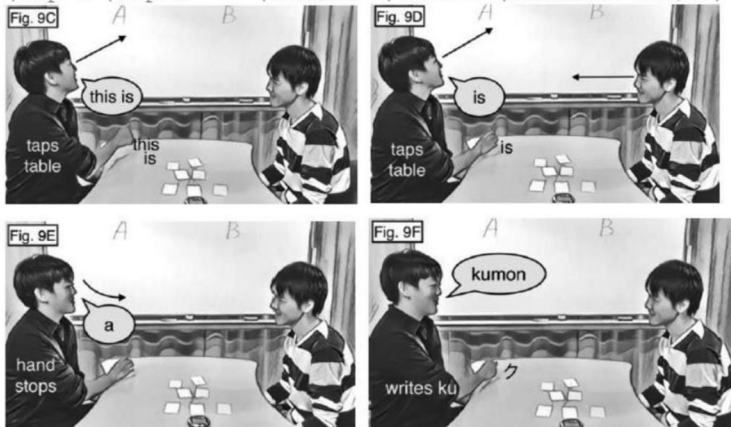
EXCERPT 7

Kumon

01 A → |I- | (0.8) |work |at |part time job
 a-gz |up/left-----|~~table/timer-->
 a-rh |touches table far left
 |rh motionless
 |touches table left
 |touches table near left
 |touches table center
 02 | (0.4)
 a-gz |~~down/left-->
 03 A |eh (0.5) |it takes |so- um | (0.9)
 fig |9A |9B
 a-gz |-----|~~up/left-->
 a-rh |raises above table/right
 |sweeps L to R
 |taps index finger
 |holds still



04 → |this |is |a(t) | (0.5) |kumon;
 |company name
 fig |9C |9D |9E |9F
 a-gz |-----|~~down/center |~~B--
 a-rh |tap L |tap C |shakes |writes (katakana ku?/ク)



05 B |°kumon | [° oh°]
 b-hd |raises brow |nods
 06 A | [°kumon°]
 a-gz |~~timer-->
 a-hd |nods

DISCUSSION

Our observations on this dataset have documented several multimodal interactional practices that novice L2 users of English incorporate into their turn construction to indicate they are visualizing syntactic elements of a turn as they produce it. The practices involve both sophisticated gaze management and precisely timed embodied representations of text via enactments of writing

or depictions of inserting words within an emergent TCU. Whether or not the speakers are actually seeing these words is not at issue: They are doing visualizing them and therefore signaling to their partner that they are working on the turn. There is emic evidence to suggest that the recipients also view it this way in that they do not begin talking while a text visualization is in progress: By not saying anything, the recipients are contributing to the accomplishment of the solitary word



search sequence. These practices therefore afford what Mushin & Pekarek Doehler (2021) called social coordination, that is, the real-time, mutually adaptive and highly situated ways in which participants synchronize their conduct.

Although these visualization practices are publicly available to both interactants, they can be directed primarily to either self or other to varying extents (see Dressel, 2020; Skogmyr Marian & Pekarek Doehler, 2022). When a speaker enacts writing in a minor way, such as by using small or truncated strokes and positioning it low on the table or on their palm (Excerpt 2), what they are writing becomes less relevant (to the recipient) than the fact that they are writing and therefore conducting a solitary word search. This is particularly the case when the speaker's gaze is directed away. However, if the speaker writes in the air with larger letters that are designed to be visible to the recipient, as A did with "marimo" in Excerpt 1, it is more clearly a case of enactment oriented to the other, such as to accomplish mutual recognition of the referent. The speaker's gaze is directed toward the recipient and the word is written within the recipient's field of vision (cf. Cibulka, 2013). The recipient treats these as designed for them by providing uptake, whereas the smaller visualization gestures do not get this sort of uptake, suggesting that the recipient also views them as primarily self-addressed.

A related but different practice involved delivering a sentence word by word by plotting each token across the table. Here, the word is represented gesturally as a unit, such as by inviting the recipient to see it positioned between the speaker's thumb and forefinger, under an open-palmed hand. Orthographic details of the word are not included as they were when a speaker air-wrote a single word. Instead, the speaker is oriented to temporally slotting out syntax. They are building a sentence, and therefore each gesture is doing more than just beating out the rhythm—the recipient is encouraged to 'see' the turn-in-progress as they hear it and to interpret the speaker as dealing carefully and deliberately with the turn construction. Further, speakers do not deploy these practices for every turn, suggesting special significance for turns in which they do. Their gaze direction displays text visualization: It is up and/or away from the recipient during the early part of the turn but shifts toward the recipient by the end. This indicates that the early part of such turns is primarily for the benefit of the speakers themselves, as they work out what to say or how to say it (forward-oriented repair or content searching). In some cases, this also led to

brief episodes of self-initiated backward-oriented repair, in which the current speaker adapted the turn on the fly, and this was reflected both in the spoken production and the speakers' pausing and revision of their embodied action.

In terms of L2 learning, these practices seem to represent outward manifestations of internal cognitive processing. While we would not be as naïve to dispute that people need to think about what they are saying—particularly when they are unfamiliar with the language they are speaking—we offer no speculation on the nature of what is happening at the neurolinguistic level during these episodes of multimodal visualization. Instead, in line with the CA approach to socially distributed cognition (Kasper, 2009), we limit our observations to externally available aspects of thought and language use, since that is all that is available to the recipient in real time as well. This is not a limitation to our study, but a strength. Any appeal to internal language processing would involve conjecture, whereas the visible and audible availability of interaction allow us as analysts to comment on behavior that is accessible to the recipient as well. Moreover, those same practices that are obtainable through microanalysis of sociality can point to the speaker's real-time orientations to emergent turn construction. Where the turn is delayed, paused, or revised, we are able to gain publicly available access to the speaker's own orientations toward language production through the interactional organization of repair.

IMPLICATIONS FOR PRACTICE

Awareness of Visual Cues to Word Searches

The data we have analyzed originate from a paired EFL speaking-proficiency test, and this peer-to-peer format has been gathering attention from assessment researchers, particularly from a CA perspective (e.g., May, 2011). Such settings can give rise to more natural interaction than that found in expert–novice tests, where the tester is often required to refrain from pursuing intersubjectivity by the test protocol (Seedhouse, 2013). The paired format also provides the test takers with the affordance of embodied interaction, including gestural orientations to text that may straddle more than one language. This poses a challenge for test raters in assessing interactional competence, and there is a need to further consider how best to incorporate such embodied practices into interactional competence rubrics (Burch & Kley, 2020; Sandlund & Greer, 2020).



In identifying and documenting the embodied practices of text visualization, this study enables such visualization to be taken into consideration in testing situations, both formal and informal. Moreover, an awareness of such practices can then be passed on to the learners themselves, and by reflecting on their purpose, students can better incorporate a range of methods for maintaining intersubjectivity and progressivity in their L2 English talk more generally.

The Intersection Between Writing and Speaking

In addition, the study has provided further insight into the relationship between spoken and written genres (Cirillo, 2019), and the analysis of these practices suggests that some novice English users are relying on their memory of written text during speaking tasks. This may be particularly the case for learners who have been taught largely through reading English rather than speaking it, as was the case for these Japanese students. It could be that pedagogical approaches that emphasize written grammar are apt to give rise to text visualization in spoken interaction, at least in the early stages of L2 learning. Inscribing a word in the air can give novice language users a chance to reflect on grammar before they use it, and slotting out sentences by visualizing their component language may help them to plan and implement a turn-in-progress while making these endeavors available to their co-interactants. These practices can also help enhance recipient comprehension by embodying the practice of turn decomposition (Svennevig, 2018). By using gestures in concert with carefully coordinated prosodic breaks, longer turns can be divided into smaller, easier-to-understand components.

Conversely, we should also consider that the visualization of text and its instantiation in particular cases can lead to a better understanding of an individual language user's current understanding of English grammar, or of the temporal unraveling of a particular turn-at-talk. This might be something that teachers and testers can make use of in assessing a student's interactional competence.

CONCLUSION

Visualizing text and using it to construct a turn is by no means a practice limited to L2 interaction—in fact, we have seen that these speakers do the same thing with lexical items from their L1 Japanese (e.g., Excerpt 1). However, since it is a phenomenon clustered around the

machinery of self-repair, it is likely that text visualization is particularly relevant for novice language users. It is apparent that careful mapping out of a sentence allows for a token-by-token examination of a turn segment as it is produced, and therefore becomes a useful affordance in moments of attention to correctness.

Although the nature of the dataset does not provide for longitudinal observations, it is worth noting that none of the cases we collected originated from what we would consider the relatively advanced English speakers in this cohort. This suggests cross-sectionally that text visualization becomes less frequent as interactional competence increases, as is the case with depictive gestures in L2 word-search sequences (Skogmyr Marian & Pekarek Doehler, 2022), and indeed with gestures in general (Eskildsen & Wagner, 2015). While additional research is needed, it is likely that as linguistic resources emerge through experience, a speaker's reliance on embodied resources lessens, providing further evidence that L2 resources evolve through and for social interaction. Alternatively, it could be that the use of text visualization practices might be one factor that makes us see its user as less proficient, since they are visibly concentrating on grammar. If so, testers could incorporate this into their assessment as one indication of beginning-level proficiency.

In addition, it is possible to view the practice as becoming established microlongitudinally (Greer, 2016; Kotilainen & Kurhila, 2020; Markee, 2011) as a part of the speaker's expanding interactional repertoire. In Excerpts 3–7, for instance, the same speaker uses text visualization to augment his turn design in multiple successive instances, and although we do not see any significant changes in the way he does this, we do find it reasonable to assume that the recipient (B) becomes increasingly aware that this practice is a regular part of this speaker's (A's) interactional repertoire. Consideration of multiple occurrences across minutes rather than months can therefore still illustrate the in situ emergence of interactional competence as a co-accomplished phenomenon.

We would suggest that our study also offers evidence of development at the turn level—not so much in terms of learning but certainly in terms of the temporal progression of utterance production. A turn that started in a rather hesitant manner becomes progressively fluent as it is mapped out across the table, or a turn that is progressing smoothly reaches a grammatical impasse without the timely availability of a lexical item that



is due, which leads to a delay in its production that is solved, in part, with an air-written gesture. Here, we are not considering the development of the speaker's interactional competence or an inventory of resources that they are able to expand, but the microdevelopment of a particular locally emergent utterance and the role that embodiment can play in achieving its production. In line with the CA approach, such a take on development focuses the observations firmly on the interaction rather than the speaker.

Even if text visualization is seen as a crutch or support, ultimately, it enables the speaker to get the turn out. It shows the recipient that the speaker is having trouble but is also dealing with it, and it generally leads to understanding. We view these microgenetic moments of intersubjective success as the building blocks for eventual development.

This study has proposed that, although publicly available, these writing gestures are often largely self-directed, as evidenced by the high speed of their production, smaller gestural strokes, averted gaze, and the fact that they are produced in a less focal space (e.g., low on the table or behind a hand). Additionally, at such times, the recipients do not visibly attend to these gestures. However, text visualizations can also be designed to be viewed by the recipient in order to display an orientation to a word as potentially out of place or requiring special attention. Either way, such writing gestures make word searches co-constructed achievements by placing them in the visible public domain. If analyzed as audio recordings alone, these moments would appear to be pauses and therefore disfluencies, but the video recordings reveal how the speakers are dealing with trouble and how the recipients are attending to these matters via gaze and by refraining from interrupting.

Our analysis, therefore, is in accord with researchers like Keevallik (2018), who see grammar as a function of not just language but the entire body. Visualizing text is designed to be seen as a medium overlaid on the talk, much like movie subtitles. It provides an additional layer to the interaction and yet is still laminated with the other modalities such that it constitutes an ensemble that exceeds the sum of its parts.

ACKNOWLEDGMENTS

This study was supported in part through JSPS Grant-in-Aid No. 17K03011. The authors are greatly indebted to the editors and reviewers of this special issue for their valuable comments.

NOTES

¹ In Sacks's (1984) seminal paper "On Doing Being Ordinary," he focuses on the public observability of members' conduct. It is not that people 'are' this or that, but rather that such states can be attributed to them by others based on what they 'do.' Similarly, here we are concerned not with participants' internal visualization, but rather how they "do visualizing" as a publicly observable practice.

² Sphere-shaped green algae (*aegagropila linnaei*), *marimo* have been the subject of several fads in Japan and are commonly sold in pet stores to be kept in small tanks similar to goldfish. *Marimo* are thus thought of as a kind of pet by Japanese people, but in a loose sense. Their "petness" is somewhat mitigated by their immobile, silent algal qualities. They are a pet you can keep even if your landlord disallows pets or if you are too busy to take care of something more substantial. In that there is no better word in English for *marimo*, A's use of the term here may also constitute a "mot juste codeswitch" (Greer, 2018).

REFERENCES

- Arano, Y. (2020). Doing reflecting: Embodied solitary confirmation of instructed enactment. *Discourse Studies*, 22, 261–290. <https://doi.org/10.1177/2F1461445620906037>
- Auer, P. (2015). The temporality of language in interaction. In A. Depperman & S. Günthner, (Eds.), *Temporality in interaction* (pp. 27–56). John Benjamins.
- Brouwer, C. E. (2003). Word searches in NNS–NS interaction: Opportunities for language learning? *Modern Language Journal*, 87, 534–545. <https://doi.org/10.1111/1540-4781.00206>
- Burch, A. R., & Kley, K. (2020). Assessing interactional competence: The role of intersubjectivity in a paired-speaking assessment task. *Papers in Language Testing and Assessment*, 9, 25–63.
- Cibulka, P. (2013). The writing hand: Some interactional workings of writing gestures in Japanese conversation. *Gesture*, 13, 166–192. <https://doi.org/10.1075/gest.13.2.03cib>
- Cirillo, L. (2019). The pragmatics of air quotes in English academic presentations. *Journal of Pragmatics*, 142, 1–15. <https://doi.org/10.1016/j.pragma.2018.12.022>
- Day, D., & Mortensen, K. (2017). Inscribed objects in professional practices: An introduction. *Journal of Applied Linguistics and Professional Practice*, 14, 119–126. <https://doi.org/10.1558/jalpp.40427>
- Deppermann, A., & Günthner, S. (Eds.). (2015). *Temporality in interaction* (Vol. 27). John Benjamins.
- Dressel, D. (2020). Multimodal word searches in collaborative storytelling: On the local mobilization and negotiation of participation. *Journal of*



- Pragmatics*, 170, 37–54. <https://doi.org/10.1016/j.pragma.2020.08.010>
- Duran, D., Kurhila, S., & Sert, O. (2019). Word search sequences in teacher-student interaction in an English as medium of instruction context. *International Journal of Bilingual Education and Bilingualism*. <https://doi.org/10.1080/13670050.2019.1703896>
- Duranti, A. (1992). Language and bodies in social space: Samoan ceremonial greetings. *American Anthropologist*, 94, 657–691. <https://doi.org/10.1525/aa.1992.94.3.02a00070>
- Eskildsen, S. W. (2018). 'We're learning a lot of new words': Encountering new L2 vocabulary outside of class. *Modern Language Journal*, 102 (Supplement 2018), 46–63. <https://doi.org/10.1111/modl.12451>
- Eskildsen, S. W., & Wagner, J. (2015). Embodied L2 construction learning. *Language Learning*, 65, 268–297. <https://doi.org/10.1111/lang.12106>
- Firth, A., & Wagner, J. (1997). On discourse, communication, and (some) fundamental concepts in SLA research. *Modern Language Journal*, 81, 285–300.
- Goodwin, C. (1994). Professional vision. *American Anthropologist*, 96, 606–633. <https://doi.org/10.1525/aa.1994.96.3.02a00100>
- Goodwin, C. (2013). The co-operative transformative organization of human action and knowledge. *Journal of Pragmatics*, 46, 8–23. <https://doi.org/10.1016/j.pragma.2012.09.003>
- Goodwin, M. H., & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, 62, 51–76. <https://doi.org/10.1515/semi.1986.62.1-2.51>
- Gough, P. B. (1972). One second of reading. *Visible Language*, 6, 291–320.
- Greer, T. (2016). Learner initiative in action: Post-expansion sequences in a novice ESL survey interview task. *Linguistics and Education*, 35, 78–87. <https://doi.org/10.1016/j.linged.2016.06.004>
- Greer, T. (2018). "And boys wore *gakuran*": *Mot juste* formulations as recalibration repair in bilingual interaction. *Japan Journal of Multilingualism and Multiculturalism*, 24, 26–47.
- Greer, T. (2019). Closing up testing: Interactional orientation to a timer during a paired EFL proficiency test. In H. T. Nguyen & T. Malabarba (Eds.), *Conversation analytic perspectives on English language learning, teaching and testing in global contexts* (pp. 159–190). Multilingual Matters.
- Hall, J. K. (2018). From L2 interactional competence to L2 interactional repertoires: Reconceptualising the objects of L2 learning. *Classroom Discourse*, 9, 25–39. <https://doi.org/10.1080/19463014.2018.1433050>
- Hasegawa, A. (2017). Collaborative orientation to the 'search for what to say' in pair work interactions. In T. Greer, M. Ishida, & Y. Tateyama (Eds.), *Interactional competence in Japanese as an additional language* (pp. 175–210). National Foreign Language Resource Center.
- Hayashi, M. (2003). Language and the body as resources for collaborative action: A study of word searches in Japanese conversation. *Research on Language and Social Interaction*, 36, 109–141. https://doi.org/10.1207/s15327973rlsi3602_2
- Heritage, J. (1984). *Garfinkel and ethnomethodology*. Polity Press.
- Holm, S., Eilertsen, T., & Price, M. C. (2015). How uncommon is tickertaping? Prevalence and characteristics of seeing the words you hear. *Cognitive Neuroscience*, 6, 89–99. <https://doi.org/10.1080/2F17588928.2015.1048209>
- Hopper, P. J. (2011). Emergent grammar and temporality in interactional linguistics. In P. Auer & S. Pfänder (Eds.), *Constructions: Emerging and emergent* (pp. 22–44). De Gruyter.
- Jefferson, G. (2004). Glossary of transcription symbols with an introduction. In G. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13–31). John Benjamins.
- Kasper, G. (2009). Locating cognition in second language interaction and learning: Inside the skull or in public view? *IRAL: International Review of Applied Linguistics in Language Teaching*, 47, 11–36. <https://doi.org/10.1515/iral.2009.002>
- Keevallik, L. (2018). What does embodied interaction tell us about grammar? *Research on Language and Social Interaction*, 51, 1–21. <https://doi.org/10.1080/08351813.2018.1413887>
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.
- Koshik, I., & Seo, M. S. (2012). Word (and other) search sequences initiated by language learners. *Text & Talk*, 32, 167–189. <https://doi.org/10.1515/text-2012-0009>
- Kotilainen, L., & Kurhila, S. (2020). Orientation to language learning over time: A case analysis on the repertoire addition of a lexical item. *Modern Language Journal*, 104, 647–661. <https://doi.org/10.1111/modl.12665>
- Markee, N. (2011). Doing, and justifying doing, avoidance. *Journal of Pragmatics*, 43, 602–615. <https://doi.org/10.1016/j.pragma.2010.09.012>
- May, L. (2011). Interactional competence in a paired speaking test: Features salient to raters. *Language Assessment Quarterly*, 8, 127–145. <https://doi.org/10.1080/15434303.2011.565845>
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.
- Mondada, L. (2014). The local constitution of multimodal resources for social interaction. *Journal of Pragmatics*, 65, 137–156. <https://doi.org/10.1016/j.pragma.2014.04.004>
- Mondada, L. (2018). Multiple temporalities of language and body in interaction: Challenges for transcribing multimodality. *Research on Language and Social Interaction*, 51, 85–106. <https://doi.org/10.1080/08351813.2018.1413878>
- Mondada, L. (2021). How early can embodied responses be? Issues in time and sequentiality.



- Discourse Processes*, 58, 397–418. <https://doi.org/10.1080/0163853X.2020.1871561>
- Mondada, L., & Svinhufvud, K. (2016). Writing-in-interaction: Studying writing as a multimodal phenomenon in social interaction. *Language and Dialogue*, 6, 1–53. <https://doi.org/10.1075/ld.6.1.01mon>
- Mushin, I., & Pekarek Doehler, S. P. (2021). Linguistic structures in social interaction: Moving temporality to the forefront of a science of language. *Interactional Linguistics*, 1, 2–32. <https://doi.org/10.1075/il.21008.mus>
- Nishizaka, A. (2006). What to learn: The embodied structure of the environment. *Research on Language and Social Interaction*, 39, 119–154. https://doi.org/10.1207/s15327973rlsi3902_1
- Pekarek Doehler, S., & Balamán, U. (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. <https://doi.org/10.1080/08351813.2021.1899710>
- Pekarek Doehler, S., & Eskildsen, S. W. (2022). Emergent L2 grammars in and for social interaction: Introduction to the special issue. *Modern Language*, 106 (Supplement 2022), 3–22. <https://doi.org/10.1111/modl.12759>
- Sacks, H. (1984). On doing 'being ordinary'. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 413–429). Cambridge University Press.
- Sacks, H. (1992). *Lectures on conversation*. Blackwell.
- Sacks, H., & Schegloff, E. A. (1979). Two preferences in the organization of reference to persons in conversation and their interaction. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 15–21). Irvington.
- Sandlund, E., & Greer, T. (2020). How do raters understand rubrics for assessing L2 interactional engagement? A comparative study of CA- and non-CA-formulated performance indicators. *Papers in Language Testing and Assessment*, 9, 132–166.
- Schegloff, E. A. (1979). The relevance of repair to syntax-for-conversation. In T. Givón (Ed.), *Discourse and syntax* (pp. 261–286). Academic Press.
- Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, 53, 361–382. <https://doi.org/10.2307/413107>
- Seedhouse, P. (2013). Oral proficiency interviews as varieties of interaction. In G. Kasper & S. Ross (Eds.), *Assessing second language pragmatics* (pp. 199–219). Palgrave Macmillan.
- Seo, M. S., & Koshik, I. (2010). A conversation analytic study of gestures that engender repair in ESL conversational tutoring. *Journal of Pragmatics*, 42, 2219–2239. <https://doi.org/10.1016/j.pragma.2010.01.021>
- Skogmyr Marian, K., & Pekarek Doehler, S. (2022). Multimodal trajectories for indexing cognitive search: Gestures in L2 word-searches and how they change over time. *Social Interaction*.
- Steinback, & Thorne, S. (2011). The social life of self-directed talk: A sequential phenomenon? In J. K. Hall, J. Hellermann, & S. Pekarek Doehler (Eds.), *L2 interactional competence and development* (pp. 66–92). Multilingual Matters.
- Stevanovic, M., & Monzoni, C. (2016). On the hierarchy of interactional resources: Embodied and verbal behavior in the management of joint activities with material objects. *Journal of Pragmatics*, 103, 15–32. <https://doi.org/10.1016/j.pragma.2016.07.004>
- Stivers, T., & Sidnell, J. (2005). Introduction: Multimodal interaction. *Semiotica*, 2005, 1–20. <https://doi.org/10.1515/semi.2005.2005.156.1>
- Streeck, J. (2008). Gesture in political communication: A case study of the democratic presidential candidates during the 2004 primary campaign. *Research on Language and Social Interaction*, 41, 154–186. <https://doi.org/10.1080/08351810802028662>
- Svennevig, J. (2018). Decomposing turns to enhance understanding by L2 speakers. *Research on Language and Social Interaction*, 51, 398–416. <https://doi.org/10.1080/08351813.2018.1524575>
- Taleghani-Nikazm, C. (2015). On multimodality and coordinated participation in second language interaction. In D. Koike & C. Blyth (Eds.), *Dialogue in multilingual and multimodal communities* (pp. 79–103). John Benjamins. <https://doi.org/10.1075/ds.27.03tal>
- Thomas, M. (2015). Air writing as a technique for the acquisition of Sino-Japanese characters by second language learners. *Language Learning*, 65, 631–659. <https://doi.org/10.1111/lang.12128>



APPENDIX

Transcription Conventions

The transcripts follow standard Jeffersonian conventions (Jefferson, 2004), with embodied elements shown via a modified version of the conventions developed by Mondada (2018).

Jeffersonian transcription symbols:

?	Rising intonation on previous syllable
ˊ	Slightly rising intonation on previous syllable
.	Falling intonation on previous syllable
,	Continuing intonation on previous syllable
↑/↑↑word	Shift to high/very high pitch
.h	In-breath
:	Prolongation of a sound
[]	Overlapped talk
w(h)ord	Laughed-through word
<u>word</u>	Stressed syllable
°word°	Audibly softer than surrounding talk
>word<	Faster than surrounding talk
(word)	Uncertain transcription
(0.2)	Silence measured in seconds
(.)	Micropause (less than 0.2 second)

The embodied elements are positioned in a series of tiers relative to the talk and rendered in gray.

	Descriptions of embodied actions are delimited between vertical bars
-->	The action described continues across subsequent lines
-->	The action reaches its conclusion
>>	The action commences prior to the excerpt
-->	The action continues after the excerpt
.....	Preparation of the action
---	The apex of the action is reached and maintained
~~~~~	Retraction of the action
~~~~~	The action moves or transforms in some way
SHIN	The current speaker is identified with capital letters

Participants carrying out embodied action are identified relative to the talk by their initial in lower case in another tier, along with one of the following codes for the action:

-gz	gaze
-lh	left hand
-rh	right hand
-bh	both hands
-px	proximity
-hd	head
-gs	gesture

Framegrabs are positioned within the transcript relative to the moment at which they were taken.

