Learner-initiated self-selection as a next speaker in a technology-mediated L2 learning environment: A multimodal conversation analytic perspective

Simin Ren [*] , Paul Seedhouse

School of Education, Communication, and Language Sciences, Newcastle University, UK

A B S T R A C T

Extensive research on next speaker selection in L2 classrooms has predominantly examined teacher-initiated nominations (e.g., Mortensen, 2008; Lauzon & Berger, 2015) or student self-selection under teacher coordination (Waring, 2011). This study shifts the focus to how L2 Chinese learners accomplish learner-initiated self-selection in a real-world, technology-mediated environment without teacher presence or institutional scaffolding. Building on Sacks et al. (1974), we reconceptualise learner-initiated self-selection as an interactional trajectory – a sequentially and multimodally achieved process, rather than a competitive act of floor-taking.

Using Multimodal Conversation Analysis (CA), we examine interactions in the Chinese Digital Kitchen (CDK), a task-based language learning environment where 72 beginner-to-advanced L2 Chinese learners cooked authentic recipes using the Linguacuisine App (Seedhouse et al., 2019). The app provided video, audio, image, and text instructions, but learners received minimal guidance and no teacher support. Analysis of the cooking sessions identifies four recurrent trajectories of learner-initiated self-selection: knowledge-display, sequential-organisation, technology-mediated opportunity, and embodied. These trajectories are not mutually exclusive but form overlapping pathways through which learners coordinate turns, manage task progression, and negotiate epistemic and procedural alignment.

Theoretically, this study contributes to CA-for-SLA by reframing self-selection as a distributed, multimodal accomplishment shaped by technological and material affordances rather than institutional regulation. It extends CA-for-SLA into non-institutional, real-world environments, showing how learners mobilise verbal, embodied, and digital resources to self-organise participation and task completion. These findings offer portable analytic categories for examining learner-initiated interaction in informal, teacher-absent, technology-mediated L2 task, and inform the design of multimodal, learner-directed learning environment.

1. Introduction

Self-selection, when a participant nominates themselves as the next speaker at a transition relevance place (TRP)—is a foundational concept in Conversation Analysis (CA), especially within the turn-taking framework established by Sacks et al. (1974). When no next speaker is selected at a TRP, any participant may self-select; typically, the earliest starter secures the turn. These principles have been

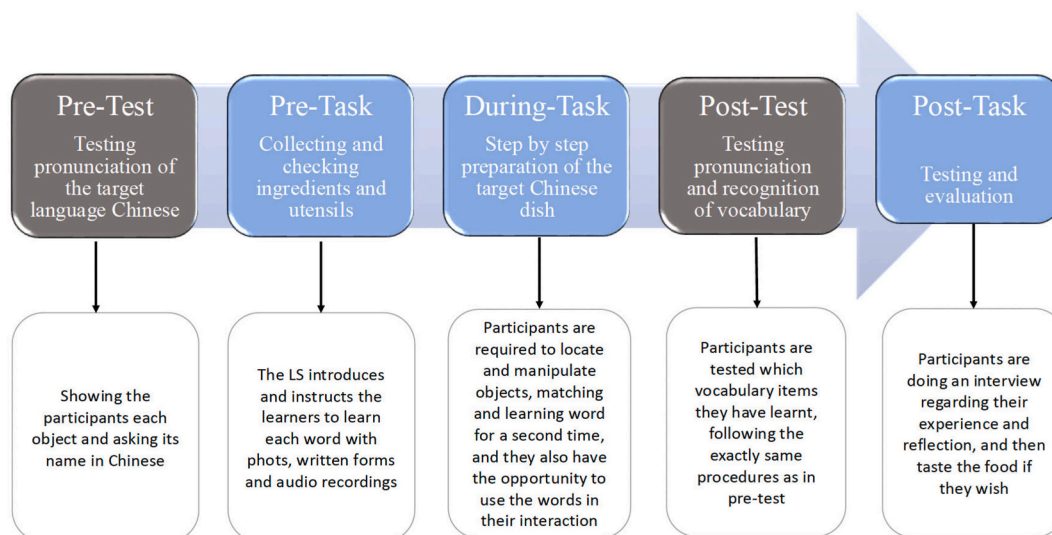| Pre-Test	Pre-Task	During-Task	Post-Test	Post-Task
Testing pronunciation of the target language Chinese	Collecting and checking ingredients and utensils	Step by step preparation of the target Chinese dish	Testing pronunciation and recognition of vocabulary	Testing and evaluation
Showing the participants each object and asking its name in Chinese	The LS introduces and instructs the learners to learn each word with phots, written forms and audio recordings	Participants are required to locate and manipulate objects, matching and learning word for a second time, and they also have the opportunity to use the words in their interaction	Participants are tested which vocabulary items they have learnt, following the exactly same procedures as in pre-test	Participants are doing an interview regarding their experience and reflection, and then taste the food if they wish

Fig. 1. Task and test cycle in the Chinese digital kitchen (CDK).

widely examined in institutional settings, particularly language classrooms where teacher-led interaction shapes participation (e.g., Hellermann, 2006; Waring, 2009). More recently, CA-for-SLA studies have begun to explore peer-led and technology-mediated interaction (e.g., Jenks, 2013; Seedhouse, 2017; Seedhouse et al., 2019), yet relatively little is known about how L2 learners accomplish speaker transitions in informal, real-world environments where no instructor, facilitator, or structured participation framework is present.

Existing research shows that learners initiate actions and manage participation in task-based or peer-led activities, often in contexts guided by teacher-designed prompts or structured materials. For instance, Hellermann (2008) traces the emergence of participation frameworks and first-turn allocations in classroom tasks, while Hasegawa (2021) demonstrates how prompts generate overlapping orientations—to completing pedagogical tasks and to contingencies of real-time interaction. These studies show that learners can and do exhibit initiative even without teachers present; however, they remain largely institution-bound, where prompts and participation structures are externally controlled. And it still lacks a fine-grained understanding of how learners organise turn-taking and role-allocation when interaction is self-directed and scaffolded only by digital prompts.

This study addresses that gap by examining learner-initiated self-selection in the Chinese Digital Kitchen (CDK), a teacher absent, technology-mediated environment where learners complete a cooking task in Chinese while receiving step-by-step instructions via the Linguacuisine app (LS1/LS2 [LS: Linguacuisine System]). The pedagogical foundation is Task-Based Language Teaching (TBLT) (Ellis, 2003; Long, 2015) (see Fig. 1), but—unlike many online or computer-assisted learning implementations—the task unfolds in a functioning kitchen, engaging multimodal, embodied resources across speech, vision, touch, and object manipulation.

The CDK presents a set of demanding interactional conditions: learners were only briefly shown the function of each app button before the during-task phase and then worked independently, encouraged to use Chinese as much as possible but without teacher/researcher presence. None of the 72 participants had used the app before, and they received no instructions on how to interact with the app or with each other. They were also informed that they could only seek the researcher's help in case of technical issues with the iPads (see Fig. 3); otherwise, they had to self-organise their interactions and resolve any issues in real time.

During the task, the Linguacuisine app provided step-by-step cooking instructions, prompting learners to perform specific actions. These instructions acted as general prompts, making both participants potential next speakers, but leaving the decision of who should respond and how entirely up to them. Consequently, learners must self-organise turn allocation, resolve uncertainties, and coordinate action in real time. In this context, self-selection becomes the key next action, signalling availability to take the turn through gaze towards the app or their co-participants often coordinating mutual orientation to initiate action. These conditions make the CDK an ideal site for investigating how self-selection, epistemic organisation, and role allocation emerge without institutional scaffolding and in coordination with digital artefacts.

Previous studies of digital kitchen projects (Kurhila & Kotilainen, 2020; Seedhouse, 2017; Seedhouse et al., 2019) have examined learning outcomes and embodied interaction but have not systematically analysed how learners manage participation and turn-taking, particularly under conditions of minimal guidance and without pre-allocation of turns. Ren and Seedhouse (2024) discuss learner-initiated testing sequences arising from self-selection between pre-task and during-task phases, but do not examine how pairs manage turn allocation as the task unfolds.

Using Multimodal Conversation Analysis (CA), we identify four recurrent interactional trajectories through which learners initiate

---

[1] Please refer to *Additional Information* for a detailed explanation of LS 1 and LS 2 in this figure.

next-speaker selection: 1) knowledge-display (displaying or claiming epistemic access), 2) sequential organisation (advancing or coordinating the procedural flow), 3) technology-mediated opportunity (aligning actions with digital prompts and interface timing), and 4) embodied (claiming turns through gesture, movement, spatial positioning). These trajectories are interactionally situated, sequentially unfolding, and often overlapping.

Theoretically, we argue that self-selection in this ecology is not primarily competitive floor-taking but an interactionally accomplished, context-sensitive practice through which learners publicly display epistemic orientation and emergent agency, while coordinating with technological affordances that shape timing and opportunities for action (e.g., pause/replay/skip). This perspective reconfigures core CA concerns—turn-taking, epistemic organisation, and role allocation—under teacher-absent, digitally scaffolded conditions, thereby extending CA-for-SLA beyond institutional, IRF-like participation frameworks.

Focusing on the during-task phase of an authentic recipe (Eggplant Stir Fry) in a student dormitory kitchen, we ask:

1. How do learners accomplish self-selection as next speaker in a teacher-absent, technology-mediated task?
2. How do these self-selection trajectories shape subsequent interactional organisation—particularly task progression, role allocation, and epistemic negotiation?

By addressing these questions, the study contributes to a broader theoretical understanding of learner initiative not as a by-product of participation, but as a central interactional mechanism that is distributed across peers, tools, and materials, and that organises the emergent turn-taking system in non-institutional environments.

2. Literature review

2.1. Turn-taking and self-selection in CA

A foundational concept in Conversation Analysis (CA) is self-selection, where speakers take the floor either through other-selection or self-selection at a transition relevance place (TRP) without being nominated (Sacks et al., 1974). In second language (L2) settings, early CA-for-SLA research often focused on teacher-fronted classroom discourse, where turn-taking is pre-allocated, and learner participation is tightly managed (Mehan, 1979; Sinclair & Coulthard, 2013; Wells, 1993; Cazden, 2001). This pre-allocation model, often realised through the IRF (Initiation-Response-Feedback) sequences, has been widely critiqued for constraining learners' interactional agency (Allington, 1980; Eder, 1982; Orletti, 1981; Lemke, 1990; Cazden, 2001; Van Lier, 2014).

Moreover, classroom interaction often unfolds as a two-party speech exchange system (SES), where the teacher often functions as the primary interlocutor, treating the class as a collective participant (Sahlström, 1999; Schegloff, 1987; Lauzon and Berger, 2015). While techniques such as traffic management (Lerner, 2003), asymmetrical responsibility (Markee, 2000; McHoul, 1978; Seedhouse, 2009; Van Lier, 2014; Walsh, 2002, 2006), and turn control reinforce this asymmetry.

Yet, studies also show that teachers can create learning opportunities through responsive turn management. For example, Sert (2017) demonstrates how teachers use embodied corrections to respond to learners' emergent knowledge displays. (Van Balen et al., 2024), drawing on Biesta's (2020) concept of 'subjectification,' argue that learners' apparent initiatives often still fall within the boundaries of teacher-controlled interaction. These insights underscore the limitations of institutional settings in fully capturing the dynamics of learner-initiated interaction.

More recent work in CA has expanded the notion of self-selection to include embodied resources and pre-emptive participation cues (Mortensen, 2008). This perspective allows for a broader understanding of turn-initiation practices that emerge outside institutional hierarchies.

2.2. Learner initiative and self-selection in peer-led and technology-mediated settings

While much CA-for-SLA research has focused on teacher-led contexts, an emerging body of work has turned to learner initiative in peer-led and technology-mediated contexts. In such contexts, self-selection does not always take the competitive from traditionally described in CA (Sacks et al., 1974), where speakers vie for the floor. Instead, learners may self-select collaboratively – through gaze, gesture, or task readiness – to initiate the next action in shared activities.

Waring (2011) defines learner initiative as any uninvited contribution, particularly when a learner takes a turn without being nominated. This includes clarification requests (Rodriguez & Wilstermann, 2018), repair initiations (Kääntä & Kasper, 2018), challenges to task procedure (Dolce & Van Compernolle, 2020), or second-position volunteer responses (Waring, 2011). These disrupt the IRF pattern (Schwab, 2011) and reflect learners' public orientation to emerging interactional contingencies.

In peer settings, Mori (2004) shows how learners negotiate procedural versus pedagogical trajectories, while Hellermann (2008) documents how task understanding is expressed through emergent turn design and embodied conduct. Hasegawa (2021) adds to this by demonstrating how learners manipulate semi-scripted prompts and physical materials to guide their turns, often adapting the task flow in the absence of external intervention.

Technology-mediated learning environments similarly open new spaces for learner initiative. Sert (2017) and Jenks (2013) demonstrate how learners use verbal and embodied resources to claim turns and manage tasks collaboratively. Koç (2023) shows how learners initiate clarification on vocabulary meaning, which contributes to situated cognition. Tai and Brandt (2018) highlight the role of teacher response to learner initiatives in multimodal classroom interaction, illustrating the imaginative potential of embodied contributions. These studies underscore the diversity of learner initiative even within institutional contexts, further highlighting the

need to investigate how such practices unfold in teacher-absent, real-world environments like the one examined in the present study.

However, many of these studies are still grounded in institutional or simi-institutional contexts. Research that systematically investigates how self-selection unfolds in non-scripted, peer-only settings – especially those shaped by digital interfaces and physical tools – remains limited.

2.3. Embodied and multimodal practices in real-world digital tasks

The digital kitchen environment offers a unique hybrid of technological and physical interaction, distinct from traditional institutional settings – providing a fertile context for exploring learner-initiated, embodied language use. Prior research has shown that learners in such environments engage in real-world, embodied tasks using technology-supported guidance. Morales (2017) illustrates how learners in the LanCook project align with app-based instructions during their cooking tasks, achieving both linguistic and procedural goals.

Kurhila and Kotilainen (2017, 2020) identify three key practices in technology-mediated peer interaction in the Finnish Digital Kitchen: 1) moving from receptive to productive language use by repeating interactional items, 1) recycling and adapting linguistic input to fit the evolving interaction, and 3) incorporating newly acquired language into real-time embodied action. Ren and Seedhouse (2024) highlight learner-initiated vocabulary testing sequences that arise prior their during-task performance and serve both epistemic and procedural functions.

These studies align with the principles of task-based language teaching (TBLT) principles (Seedhouse, 2017), showing how learners engage with real-world, authentic cooking task, involving actual ingredients and tools, real cooking, and environmental contingencies, performed in a functioning domestic kitchen through peer coordination, rather than simulated or scripted classroom activities/non-simulated cooking tasks. Yet, they do not systematically examine the sequential organisation of turn-taking. Mortensen (2009) and Hasegawa (2021) argue that in environments where no external turn allocator exists, learners rely on embodied cues and material affordances to negotiate speaker roles.

In the CDK, unique contingencies arise: the app does not nominate speakers, and learners receive no guidance on interactional roles. Some treat the app like a teacher, awaiting prompts, while others self-select freely, replaying instructions as needed. While learner agency and multimodal engagement have been well-documented, the interactional mechanics of turn-taking remain underexamined.

2.4. Positioning the present study

This study builds on and extends the above strands of research by analysing learner-initiated self-selection in an unscripted, embodied, and digitally mediated cooking task. By shifting analytic attention to moment-by-moment, multimodal coordination, we reconceptualise self-selection not as a competition for floor space but as a collaborative, emergent, embodied task-driven achievement, where learners must manage interactional trajectories autonomously, orienting to digital prompts and embodied contingencies without external facilitation.

Through a close analysis of interaction in the CDK, we aim to understand how learners manage speaker transitions without institutional scaffolding. In doing so, this study contributes to CA-for-SLA research by examining how embodied action, peer coordination, and technological prompts jointly shape turn-taking trajectories to shared tasks and technologies in real-world learning environments.

3. Research design

We investigate how L2 Chinese learners manage turn allocation in the absence of a language teacher. Specifically, learners in the CDK must determine how to interact with each other and the digital app, decide when and how to respond to cooking prompts and learning resources provided by the app, and figure out how to use the app and the environment effectively and adaptively to meet their own needs. Our analysis focuses on the during-task phase, examining participants' situated orientations to the interactional affordances of the environment as they become relevant in interaction. The following research questions guide the study:

1. How do learners accomplish self-selection as next speaker in a teacher-absent, technology-mediated task?
2. How do these self-selection trajectories shape subsequent interactional organisation – particularly task progression, role allocation and epistemic negotiation?

To address these questions, we employ multimodal Conversation Analysis (CA) to examine how learners organise interaction and manage turn allocation. Situated within applied CA (Ten Have, 2007), this study focuses on participant conduct in naturally occurring, institutionally relevant settings. Since the 1990s, CA has become increasingly prominent in L2 research (Firth & Wagner, 1997; Gardner, 2004; Kasper, 2006, 2009; Mori & Zuengler, 2008; Seedhouse, 2009) for its capacity to capture the complexity of learning interactions.

Previous CA studies have shown the emergent and dynamic nature of learning interactions (Frazier, 2007; He, 2000; Hellermann, 2006; Kasper, 2006; Mori, 2004, 2007), documented learning trajectories over time (Brouwer & Wagner, 2004; Hellermann, 2006, 2008; Markee & Kasper, 2004; Young & Miller, 2004), and analysed interactional practices that shape participation opportunities through sociocultural lens (He, 2004; Pavlenko & Lantolf, 2000; Waring, 2008).

Drawing on this tradition, our analysis aims to capture learner-initiated self-selection practices, focusing on the design, sequential organisation, and multimodal composition of initiative turns.

The dataset consists of 36 video recordings of CDK participants completing a cooking task. These recordings were transcribed using Jefferson's (2008) system for verbal data and Mondada's (2018) multimodal transcription conventions to capture embodied actions (see Appendix A). Pinyin was added to reflect Chinese-language use in interaction.

4. Analytical procedure and collection formation

This study adopts a collection-based Conversation Analysis approach (Schegloff, 1987; Sidnell, 2010), focusing on self-selection as next speaker. The dataset comprises 42 h of video-recorded interaction from 32 learner pairs completing the CDK tasks, with each pair recorded for approximately 38 min on average.

Following full transcription of the during-task phase using Jeffersonian (2008) and Mondada's (2018) conventions via Transana software, we identified instances where learners initiated turns without explicit selection, typically in response to app prompts or peer cues. Self-selection was defined as any learner-initiated turn occurring without prior speaker allocation, generally at or just before a transition relevance place (TRP).

From this, we complied 36 cases for analysis, using iterative, comparative methods to identify recurrent interactional patterns. This yielded four recurrent trajectories of self-selection: 1) knowledge-display; 2) sequential organisation; 3) technology-mediated opportunity; 4) embodied.

Each trajectory was grounded in both sequential organisation and multimodal composition. Excerpts were selected based on clarity, typicality, and analytic relevance. The analysis was collaborative and recursive, involving repeated viewings, discussion among the authors, and reflexive checking to ensure analytic consistency and robustness.

5. The Research Setting and sample

The learners in this study are 6 groups of L2 Chinese learners at beginners, intermediate and advanced levels, studying in China. They are situated in the kitchen of their international students' dormitory, where their language learning task is to prepare a Chinese dish ('*Eggplant Stir Fry*') using the prepared cooking ingredients and utensils by the researcher. In addition to the cooking task, they are also learning a related set of L2 vocabulary. This takes place in a real-world, non-virtual environment, where the leaners cook actual food in a real kitchen. However, it is a digitally-mediated environment, as the learners must interact with a digital system (the Linguacuisine app) to complete the task. Seedhouse (2017) refers to this a real-world digital environment. The learners receive on-demand instructions and assistance in L2 Chinese via video, audio, photos, and text from a tablet, as shown in Fig. 2 below.

Fig. 2 presents a typical screenshot from the tablet that displays the cooking instructions available to the learners. A researcher is present during the task to record the interaction but does not provide any assistance or feedback. As mentioned earlier, the learners have not received any explicit guidance on how to interact with each other or the tablet; they have only been shown which buttons to press on the tablet to access help in the form of video, audio, text, and photos related to the recipe.

As illustrated in Fig. 3, data were collected in the communal kitchen of the participants' dormitory at three universities in Xi'an, China. The cohort consisted of 72 international L2 Chinese students, including 43 males and 29 females, aged between 18 and 40 years old. Their exposure to Chinese ranged from 2 months to 68 months, with a mean of 13 months (just over one year). Participants were typically paired so that one had a higher level of language proficiency than the other. In most cases, participants did not share a common L1 and used Chinese or English as their L2 throughout the task. In instances where participants shared a common L1 (e.g.,

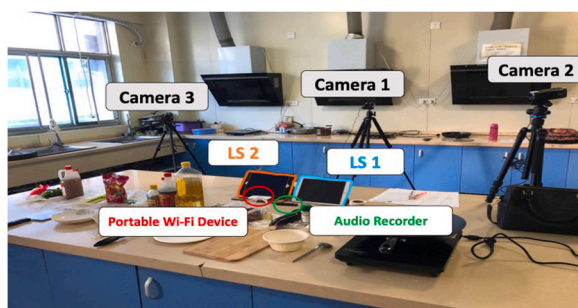**Fig. 2.** View of linguacuisine app tablet screen with Chinese instructions.

Fig. 3. Research setting.[1]

Extracts 1, 5, 6), they were asked to speak in either Chinese or English L2, although there were occasional instances of code-switching between the two languages.

6. Data presentation

As outlined in the Introduction, CDK participants faced various challenges that introduced uncertainties into the interactional environment, shaping their decisions regarding when and how to initiate self-selection. This section presents illustrative extracts that exemplify four primary trajectories of self-selection observed in the CDK context: knowledge-display, sequential organisation, technology-mediated opportunity, and embodied. These examples serve to illuminate the underlying motivations for self-selection and the specific interactional practices through which participants enact it.

In all transcripts, we adopt Mondada's (2018) definition of *gaze* as a methodical, temporally unfolding embodied action that works in coordination with talk and other bodily behaviours to organise interaction. Accordingly, in this study, *gaze* refers to a sustained visual orientation toward a target (e.g., a person, object, or screen in our context), often lasting several seconds and functioning as part of participants' embodied interactional strategies in the CDK.

6.1. How learners accomplish self-selection: multimodal and sequential trajectories (RQ 1)

This section presents four recurrent interactional trajectories through which learners accomplished self-selection as next speaker in the CDK environment. These are not discrete or mutually exclusive types but represent dominant interactional configurations – grounded in participants' orientations to epistemic access, task progression, technological affordances, and embodied coordination. In all cases, self-selection is achieved through the strategic deployment of multimodal resources (e.g., gaze, gesture, object handling, verbal repetition) and interactional practices (e.g., pre-positioning, turn anticipation, post-completion entry) that jointly constitute an emergent, distributed turn-taking system in the absence of institutional scaffolding.

Each analytic trajectory below is framed through a common lens: the sequential environment in which self-selection arises, and the multimodal composition of the learner's turn-initiation. These trajectories collectively reveal how learners manage speaker transitions interactionally – not merely when or why they self-select, but through what observable practices and resources self-selection is constructed.

6.1.1. Knowledge-display self-selection: claiming the floor via epistemic assertions

This section shows how learners self-select by displaying relative epistemic access: using gestures, gaze, and verbal repetition to claim the turn at or near transition-relevance places (TRPs).

In these cases, learners initiate turns that index either knowledge possession (K+) or knowledge gap (K-), positioning themselves as the next relevant speaker through verbal repetition of key vocabulary, pointing to the referents, or seeking confirmation. These epistemically framed actions often occur alongside the app's instruction delivery or in its immediate aftermath, where the task requires object identification or instruction interpretation.

6.1.1.1. K + stance knowledge assertions. In the following *Extract 1*, illustrating K+ self-selection, two intermediate level participants, HIL and DEL, are working on step 12. The target ingredient for this extract is ***'sheng chou'*** (soy sauce). HIL's language proficiency level is slightly higher (HSK 4) than DEL (HSK 3).

‡ HIL ‰ DEL

	HIL	DEL
HSK Level	HSK 4 (intermediate)	HSK 3 (intermediate)
Ethnicity	Norwegian	Norwegian
Gender	Female	Female
Age Group	18-25	18-25
University Programme	Chinese Language and Literature (UG)	Chinese Language and Literature (UG)
Arrived in China	09/2016	09/2016
Length of Stay in China	2 years and 6 months	2 years and 6 months

```
339          ‡(0.4) ‡

      hil    ‡leans forward and presses the play button on LS 1‡

340   LS 1   ‡#xian zai, qing zai bai se de da wan li ‰# yi ci dao ru shi liang de

             Now, in the small white bowl,  please add in order, the appropriate amount of

 →    hil    ‡leans forward and gazes at LS 1------------------------------->>

      del                                        ‰RH points to ingredient lists on

                                                  LS 2 and gazes at LS 2---->>

      ls 1   Video Demonstration on LS 1 starts---------->>

      ls 2                                      Ingredient List on the APP (LS 2)

      fig       #fig.1                                    #fig.2
```
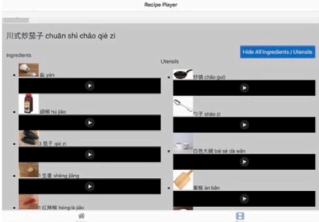

Extract 1

1a

2

2a

341	LS 1	*#sheng chou*
		soy sauce
fig		#fig. 3

3

342→	DEL	‰ *#[sheng chou]* ‰
		[soy sauce]
Del		‰points at the
		'soy sauce' entry

. (*continued*).

```
fig      on LS 2 (生抽 shēng chōu)------->>

          #fig. 4
```


4

4a

```
          (A sample picture of the target object, audio recording of target
          word pronunciation, Chinese Character and Pinyin)
```

343→ **HIL**   *##[sheng chou], °zhe ge°##*

  [soy sauce], ᵒthis one ᵒ

hil   **⧺RH on the soy sauce container⧺**

  #fig.5

5

344   **DEL**   *⧺*‰ *en* ⧺*‰*

       em-hmm

del   **⧺gazes at**

  LS 1--->⧺

hil   **‰gazes at**

  soy sauce‰

fig   #fig.6

6

345   **LS 1**   *hong tang*

  brown sugar

                                    . (*continued*).

In this interaction, HIL accomplishes self-selection by displaying epistemic access to the referent *'sheng chou'* (soy sauce), thereby claiming next speakership in the unfolding interaction. Following LS 1's instruction (line 340–341), HIL enters the turn space without pause, coupling verbal repetition ('*sheng chou*') with a deictic gesture ('*zhe ge*' – this one) directed at the physical bottle (line 343, fig. 5). The self-selection is multimodally composed: it is preceded by gaze anchoring on the app screen (line 340, figs. 1 and 2), then synchronised with bodily orientation and referential action. This turn emerges immediately after the app's verbal cue – a transition-relevance place in which the next speaker has not been re-assigned.

DEL, who initiates a parallel pointing gesture to the same referent via LS 2 (line 342, fig. 4), subsequently ratifies HIL's selection with an embodied minimal response ('*en*') and convergent gaze on the physical object (line 344). This co-participant alignment confirms HIL's epistemic display and locks in the referent, completing the sequence without conflict or delay.

This extract shows how learners use multimodal resources to project, claim, and ratify next speakership. HIL's self-selection is achieved not through explicit nomination or turn-competition, but through interactionally visible epistemic positioning that verbalised and embodied in synchrony with sequential cues. The turn is structurally complete, epistemically assertive, and socially displayed, demonstrating how self-selection is achieved through the coordinated use of verbal repetition, gaze, and embodied reference within an emerging turn-taking environment.

6.1.1.2. K- stance knowledge assertions. In *Extract 2*, by contrast, two advanced-level participants, LI and YING, demonstrates K- self-selection after failing in their initial attempt to locate the same object as in *Extract 1* – *'sheng chou'* (soy sauce).

```
+ Li  ∉ YING
```


	LI	YING
HSK Level	HSK 5 (advanced)	HSK 5 (advanced)
Ethnicity	South Korean	Kyrgyzstani
Gender	Female	Female
Age Group	18-25	18-25
University Programme	Business (MA)	Business (MA)
Arrived in China	2017	10/2016
Length of Stay in China	2 years	2.5 years

```
112          (0.3)

113   LS 1   +*∉xian zai, qing zai bai se de da wan li yi ci dao ru shi liang de

             Now, in the small white bowl, please add in order, the appropriate amount of

      li     +starts gazing at the LS 1--------------------------------------------------------->>

      ying   ∉starts gazing at the LS 1--------------------------------->>

114   LS 1   =sheng chou (0.3) hong tang (0.3) yu mi dian fen

             =soy sauce (0.3) brown sugar (0.3) corn starch

115→  YING   ∉ # sheng chou shi shen me ya?

             what is soy sauce

      ying   ∉LH starts searching on the LS 2

             (ingredient list)

      fig.     #fig. 7
```



```
116   LS 1   # zhi ma you

             sesame oil

      fig    #fig. 8
```
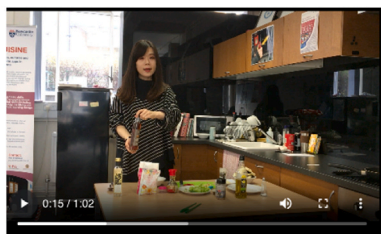
Extract 2

8

117→   LI   = +# zhi ma you +

         =sesame oil

li   +LH on the sesame oil bottle +

fig   #fig. 9

9

118   LS 1   =he mi cu

         =and rice vinegar

119→   LI   sheng chou ying gai shi zhe ge ba, +(0.2) CHOU +

         this should be soy sauce (0.2) CHOU

li   +LH on soy sauce------------------>>

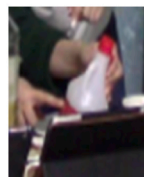   +points at the written word

                                            'chou'(抽) on real object +

fig   #fig.10

10

10a

10b

. (*continued*).

After LS 1 lists ingredients (line 114), YING initiates a repair (line 115) with a *wh*-question (Raymond & Heritage, 2013), '**sheng chou shi shen me ya?**' (What is soy sauce?), entering the turn space while LS 1 is still delivering the instruction. This move performs two sequential functions: 1) it projects an epistemic lack that requires uptake; 2) it creates an actionable next-speaker slot by publicly indexing a knowledge problem. YING's embodied conduct (searching LS 2, line 115, fig. 7) accompanies the verbal query, making the epistemic trouble visible and recognisable to LI.

More interestingly, LI chooses not to step in to address YING's epistemic gap until line 119, after repeating LS 1's last object

instruction ('**sesame oil**') as a form of acknowledgment and then she orients back to LI's original query. In line 119, LI uses a combination of verbal explanation ('**sheng chou ying gai shi zhe ge**' – this one should be soy sauce) and a deictic point to the soy sauce object, and a point to the written Chinese character ('chou') on the container (figs. 10 and 10a). These resources jointly instantiate a link between word and referent, thereby closing the repair sequence.

We also observe notable differences between LI's and YING's approaches to self-selection. LI initiates her self-selection without waiting for LS 1 to complete its turn, while YING chooses to wait for the completion of LS 1's turn. In other words, LI prioritises addressing her own epistemic issue without considering LS 1's turn continuity, anticipating that a potential turn relevance place (TRP) could emerge at any point during LS 1's delivery for a self-selection. However, YING acknowledges the completion of LS 1's turn as the appropriate TRP before self-selecting or orienting to her co-participants LI's epistemic query.

Unlike the symmetrical epistemic alignment in *Extract 1*, *Extract 2* reveals asymmetry, where the less knowledgeable participant (YING) relies on the K+ participant (LI) to address her epistemic gap. Although the technology (LS 1/LS 2) provides contextual cues, it does not directly respond to learner queries unless actively engaged, creating space for knowledge negotiation.

In both cases, learners accomplish self-selection by indexing their stance toward the knowledge in play. Crucially, these are not only knowledge claims but interactional moves that achieved through fine-grained multimodal coordination that renders epistemic positioning publicly visible and sequentially relevant. Gaze direction, object manipulation, and verbal repetition jointly function to produce a recognisable self-selection slot within an emergent turn-taking.

6.1.2. Sequential-organisation self-selection: advancing task trajectories

This section explores the sequential organisation trajectory, demonstrating how learners self-select to manage or advance the sequential organisation of the task: often pre-empting silence, prompting next actions, or coordinating procedural steps through embodied or verbal cues.

6.1.2.1. Role-negotiation and task advancement. In *Extract 3*, two intermediate learners, LING and LI, demonstrate how self-selection is used to assign roles as they prepare to begin the *during-task* phase. LING, whose Chinese proficiency is slightly higher (HSK 4) than LI (HSK 3), initiates this process. Through self-selection, they collaboratively negotiate and allocate responsibilities, facilitating coordinated task engagement.

	LING	LI
HSK Level	HSK 4 (Intermediate)	HSK 3 (Intermediate)
Ethnicity	Japanese	Kazakhstani
Gender	Female	Female
Age Group	18-25	18-25
University Programmes	International Higher Education (UG)	International Higher Education (UG)
Arrived in China	09/2017	09/2017
Length of Stay in China	1.5 years	1.5 years

```
501→  LI      ex::

502→  LING    ∝ * ∂ # xian kan zhe ge? ∝ * ∂
               look this one first?

      ling    ∝ points at the LS 1, gazes
               at LI------------------>∝

      li      ∂ gazes at LING-------->>∂

      fig     #fig.11
```



```
503           (0.5)
504   LI      dui, ∂ hehehehehe ∂
               yes, .hhhhh

      li      ∂ RH points to the
               LS 1 and clicks
               the play button ∂

505   LS 1    ∝ * ∂ qing ba san ge # qie zi dou qie cheng liang ban
               please chop the three eggplants into halves

      ling    ∝ gazes at the LS 1------------------------>>

      li      ∂ gazes at the LS 1----------------------->>

      ls 1                        #(in LS 1 video demonstration) picks up
                                  the eggplant---------------->

      fig                        #fig. 12
```



```
506           ∂ #(0.2)
   →  li      ∂ LH picks up
               one eggplant
               from her left
               side and continue
               gazing----->>

      fig     #fig.13
```

Extract 3

13

```
507            #(10.4)

     ls 1   LS 1's non-verbal demonstration------>>

            (both participants continue gazing at it)

     fig    #fig.14
```


14

```
508            ∂ (0.6) ∂

     li     ∂ stops the LS 1

            and leans right

            to LING ∂
```
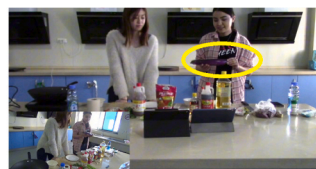
```
509→   LI     ∂ er:: xian zai:: wo men yao:: ∂ # QIE ZI! ∂

              er:: now we need: EGGPLANT!

     li     ∂ both hands hold the eggplant

            and gazes at it------------->>∂

                                    ∂RH holds the eggplant

                                    and gazes at LING∂

     fig    #fig.15                        #fig.16
```
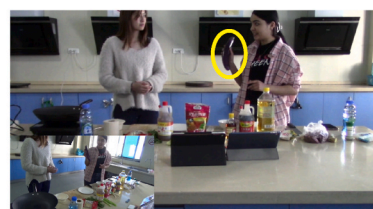

15

16

. (*continued*).

```
510    LING    hao de
               ok

511→   LI      ∂ # ha ha ha (0.8) ∂ wo gei ni
                                           I give you

       li      ∂ puts the eggplant
                 on the chopping
                 board-------->>∂

       fig     #fig.17
```



```
512    LING    ∂ ha ha ha hao de ∂
                        ok

       li      ∂moves to her left and
               picks up the kitchen
               knife--------->>∂

513            ∂ (0.2) ∂

       li      ∂gives the
               kitchen knife
               to LING∂

514    LING    ∝ xie xie ∝
                  thank you

       ling    ∝ takes the
               knife from LI∝

515    ling    ling starts to chop the eggplant------>>
```

. (*continued*).

14

As the outset of the task phase, LI's hesitation token ('*er::*' in line 501) opens a space of uncertainty. LING self-selects in line 502 to manage this gap, proposing '*xian kan zhe ge?*' (look at this one first?) while simultaneously pointing to the LS 1 and gazing at LI (fig. 11). The turn combines a directive proposal with an embodied indexical cue, orienting both participants toward the instructional resource. LI aligns by agreeing verbally ('*dui*' – yes) and physically enacting the instruction – pressing play on LS 1 (line 504).

As the app produces the instruction (please chop the three eggplants into halves, line 505), both participants sustain gaze on the video demonstration (fig. 13). LI's next action (line 506) involves picking up the eggplant, mirroring the on-screen action. Despite the extended demonstration (lines 507 and 508), self-selection is withheld until LS 1's turn closure. Then, LI initiates speaker transition (line 509) through a composite multimodal turn: she announces the next step ("*wo men yao :: qie zi*"- we need::eggplant), holds up the target object, and orients her gaze toward LING (figs. 15 and 16). This self-selection displays actional clarity and re-anchors joint attention onto the task material.

The next self-selection occurs in line 511, marking a pivotal role transition. Before this, in line 510, LING agrees with LI's verbal announcement while performing a 'chopping' gesture with her right hand. This gesture not only reinforces the task instruction but also signals to LI that the next action involves chopping the eggplant. Simultaneously, LING steps back, offering LI the physical space to carry out the step, implicitly acknowledging LI's leadership role due to her active engagement and relatively higher proficiency.

What follows in is a role negotiation sequence. In line 511, LI places the eggplant on the chopping board and reallocates the task to LING, stating, '*wo gei ni*' (I give you). Marks the physical and verbal transition of responsibility. LING then verbally affirms (lines 512), then takes up the chopping tool and receives the eggplant from LI, indicating task handover. This is not a unilateral assignment but a collaboratively ratified transition, evidenced by LING's uptake ('*xie xie*' – thank you) and immediate initiation of the next embodied task action (line 515).

The self-selection we discussed in this extract functions as: 1) Resolve initiation gaps; 2) Coordinate shared attention; 3) Advance the instructional sequence; and 4) Negotiate and redistribute task roles. Rather than relying on explicit turn allocation or verbal nomination, the participants use gaze, object manipulation, and embodied directives to manage turn transitions. This demonstrates how self-selection supports smooth, task-relevant progression within an emergent sequential structure.

We see in the following *Extract 4* for LI and LING's step 2 and we see in line 518 that LING initiates '*na ni::*' ('then you:: (do this step)'). The way they organise their during-task phase is one complete one step and then they switch in a collaborative and harmonious way.

```
515    LS 1    ran hou, zai ba mei yi ban qie cheng xi tiao zhuang

               Then, chop each half into strips

516            (16.1)

               (LS 1's non-verbal demonstration of step 2)

517    LI      hao de

               ok

518→   LING    na ni::

               Then you::

519    LI      =en! ha ha ha

               =ok/of course!
```

Extract 4

6.1.2.2. Sequential reorientation through self-selection. Extract 5 illustrates how a participant reclaims trajectory control through multimodal self-selection that prioritises collective task progression over individual queries.

In this example, two participants, DEL and HIL, are engaged in step 12 of the recipe. Prior to line 366, DEL initiated a sequence by asking for the English names of the ingredients listed in this step, including soy sauce, brown sugar, sesame oil, rice vinegar, and corn starch.

`HIL‡  DEL ‰`

	HIL	DEL
HSK Level	HSK 4 (intermediate)	HSK 3 (intermediate)
Ethnicity	Norwegian	Norwegian
Gender	Female	Female
Age Group	18-25	18-25
University Programme	Chinese Language and Literature (UG)	Chinese Language and Literature (UG)
Arrived in China	09/2016	09/2016
Length of Stay in China	2 years and 6 months	2 years and 6 months

```
366        (0.2)

367→   DEL  ‰ # zhe  [ge  ne?]‰

               [this one]

   del   ‰points at the corn

         starch, gaze moves

         back to the LS 1 ‰

   fig     #fig.18
```



```
                                    18
```



```
                                    18a
```

```
368→   HIL        ‡# [fang duo shao] ‡

               [how much shall we put in]

   hil        ‡gaze moves back to

              the LS 1---------->‡

   fig         #fig.19
```

Extract 5

19

```
369    DEL   ‰*‡ # .he he he £ ‰*‡

       del   ‰laughs and moves gaze

             to the LS 1------>‰

       hil     ‡smiles---------->‡

               #fig.20
```


20

```
370    HIL   ‡*‰ #.he he he £ ‡*‰

371→   HIL   ‡ # zai, zai ‡ # kan (0.2) yi bian

                 again, watch again (0.2) one more time

       hil   ‡RH points at LS 1‡

                             ‡ leans across and

                               rewinds the video------->>

       fig     #fig.21          #fig.22
```


21

22

. (*continued*).

In the initial sequence, DEL attempts to identify the English name of a target object, the corn starch, using a deictic gesture (fig. 18 and 18a) and a verbal query ('***zhe ge ne***?' (this one?)). Her gesture of pointing to the corn starch, combined with a brief gaze shift to the LS 1, indicates her intent to integrate both visual and verbal resources for clarification. However, this potential uptake slot is not ratified. HIL instead immediately self-selects (line 368) with '***fang duo shao***' (*how much shall we put in?*), redirecting attention toward quantity rather than identity, aligning her gaze with LS 1 (line 368, fig. 19). This move resists uptake of DEL's prompt and asserts a completing sequential trajectory: one that aligns with progressing the cooking step.

This self-selection is multimodally reinforced. HIL's gaze move back to LS 1 (fig. 19), redirecting interaction around the visual instructional source. The paired verbal and embodied turn functions as a task-oriented pivot, prioritising progression over individual lexical clarification. DEL's laughter (line 369) and shared gaze shift (fig. 20) suggest momentary alignment, not insistence. HIL's second self-selection (line 371) closes the sequence: '***zai kan yi bian***' (watch again one more time), followed by reaching over to rewind the video (figs. 21–22). This physical manipulation of the shared technological resource both confirms her trajectory control and makes turn priority visible. Her pointing gesture and physical manipulation of the LS 1 especially leans across DEL suggests her initiative in redirecting the interaction toward the primary task. In other words, in these language learning sequences initiated by DEL, DEL is in fact nominate HIL as the next speaker in line 367, however this nomination is resisted by HIL.

This extract illustrates sequential dominance enacted through self-selection. Unlike role assignment (*Extracts* 3–4), here self-selection is used to suspend 'parallel activity' (Koole, 2007) and enforce task linearity. DEL's initial move treats HIL as a next speaker by presenting a question, but HIL counters this by assuming speaker role herself and redefining relevance. Additionally, unlike the previous interactions, where technology serves as a supportive or secondary resource, here LS 1 is framed as a central authority that participants rely on to guide task progression, therefore shape and results in HIL's self-selection in line 368 and 371. The negotiation between DEL and HIL demonstrates how learners balance their individual learning objectives with the shared responsibility of task completion, using multimodal resources such as gestures, gaze, and embodied actions to navigate these competing demands.

Here, self-selection is not about claiming knowledge, turn-filing, or repair-initiation, but about reorientation device to regulate task order, maintain instructional pacing, and suppress divergence, even in peer-only settings without teacher scaffolding. Learner's claim turns not to resolve uncertainty but to maintain coherence, avoid drift, or ensure the task proceeds. These self-selections are enacted through directive formats (e.g., '*then you do ….*'), bodily repositioning, and handling of tools – actions that render the next speaker slot socially recognisable in the absence of pre-assigned roles.

6.1.3. Technology-mediated opportunity for self-selection: temporal claiming via interface control

This section examines how learners self-select by aligning with the temporal affordances of the Linguacuisine App – pausing, replaying, or anticipating prompts as a means to claim next speakership. In particular, how touchscreen engagement is used to align turn-taking with interface temporality, positioning the app as a cueing mechanism rather than an instructive authority.

In contrast to previous sections, the focus here is not on epistemic stance or procedural advancement per se, but on how learners interact with technology as part of their turn-construction. Self-selection occurs through strategic manipulation of the app's interface, used to create or withhold opportunities for speaker transition. Participants treat the app's timeline as a shared interactional resource, cuing themselves and each other into next actions.

Unlike other groups, the pair in *Extract 6* had already watched the entire during-task process (31 steps) before beginning their own cooking phase. As a result, the focus here shifts from role assignment to the use of technology as a prompt rather than as a primary source of information. The extract opens with YEDA confirming task readiness with the researcher, who was simultaneously checking the cameras. Following this brief exchange, the pair proceeds to step 1 of their during-task phase.

```
YEDA +   SHASHA ±
```


	SHASHA	YEDA
HSK Level	HSK 4 (intermediate)	HSK 3 (intermediate)
Ethnicity	Tajikistani	Kazakhstani
Gender	Female	Female
Age Group	18-25	18-25
University Programme	Cross Cultural Communication (MA)	Cross Cultural Communication (MA)
Arrived in China	09/2017	09/2017
Length of Stay in China	2.5 years	2.5 years

```
399    YEDA      + ke yi kai shi ma? +

                 Shall we start?

                 +looks at researcher+

400    RES       ± ke yi kai shi ±

                  yes you can start

shasha      ±moves to her left,

            eye gaze on the

            chopping board ±

401→   YEDA      +*± # di yi ge + shi:: (0.3) an ban

                 The first one is:: (0.3) chopping board

yeda      +turns gaze from

          LS 1 to SHASHA +

                    +right hand points at the chopping board

shasha      ±both hands pick up the chopping board, looks at

            YEDA and smiles----------->>

fig       #fig.23
```


23

Extract 6

```
402              +*±# (1.4) +*±
→      yeda      +leans forward
                 and presses the
                 play button
                 on the LS 1+
       shasha     ±puts down the
                  chopping board
                  on the table,
                  eye gaze turns
                  to the LS 1 ±
       fig         #fig.24
```


24

24a

```
403    LS 1    qing ba san ge [qie zi]
                please use three [eggplants]
       ls 1    *verbalise  the  instruction  and  then  starts  its  non-verbal
                demonstration
404→   YEDA                    +#[en::] qie zi!
                                  [em::] eggplant!
       yeda                    +leans forward and pauses the LS 1 with RH
       fig                      #fig.25
```


25

```
405→   SHASHA   ± qie zi shi:: ±# °zhe ge ma°? ±
                   eggplant is::°this one°?
       shasha   ±moves to her right±
                            ±RH on eggplants
                            and looks at YEDA±
       fig                  #fig.26
```

. (*continued*).

20

26

```
406    YEDA     +# en? +

       yeda     +points

                to SHASHA

                and smiles+

       fig      #fig.27
```


27

```
407    SHASHA   yi ge qie zi ma?

                One eggplant?

408    YEDA     em:: dou ke yi

                em:: I don't mind

409    shasha   ±picks up the kitchen knife and starts chopping----------->>
```
. (*continued*).

To begin the cooking task phase in line 401, YEDA self-selects as the first speaker, initiating step 1 with the instruction, '***di yi ge shi::***' (the first one is::) The elongation of '***shi::***' (is::) suggests either she is recalling the first step or the target vocabulary they are expected to use. This utterance operates in synchrony with her embodied action, gaze shift and pointing, directed at the chopping board (fig. 23), thus constituting a multimodal pre-positioning for self-selection. Her turn is not just verbal: it is tactically enacted in spatial deixis and semiotic readiness.

SHASHA responds not with speech but with action: she picks up the object and meets YEDA's gaze. This response, temporally and physically aligned, validates YEDA's selection and ratifies the turn without repair or contestation. Notably, SHASHA's embodied compliance happens before any explicit instruction from LS 1 has played, foregrounding peer coordination over system-following.

YEDA responds to SHASHA's action by verbalising the name of the object in Chinese, '***an ban***' (chopping board), while simultaneously pointing to it (line 401, fig. 23). This moment shows SHASHA visually attending and redirecting her utterance to YEDA and then physically engaging with the object, reinforcing their collaborative approach. SHASHA's smile further strengthens their mutual understanding and coordination.

YEDA then physically initiates the app's playback (line 402, fig. 24 and a). This touchscreen act is both an instruction-following move and a device-mediated self-selection, controlling when the instructional turn begins. Importantly, she interrupts the LS 1 shortly thereafter in line 404 by pausing it as soon as the word '***qie zi***' (eggplant) is spoken. The pause and verbal echo ('***en:: qie zi***') function as lexical anchoring, asserting control over when to process and when to delay progression. Here, touchscreen pausing replaces traditional verbal floor holding – the learner uses the device as a regulator of sequential space.

This analysis shifts emphasis from task phase management to interface-temporal management. YEDA's actions are less about role delegation and more about tempo control: pacing the interaction to match their mutual readiness. SHASHA's turn in line 405 ('***qie zi shi zhe ge ma?***' – eggplant is this one?) is produced under this pause condition, indicating that YEDA's intervention has created a relevant space for peer clarification.

Throughout the interaction, YEDA and her self-selection play a critical role in coordinating both the verbal and non-verbal aspects of the task, as seen in lines 401 and 404. Her active participation demonstrates her leadership in managing the task and synchronising it with LS 1's instructions, though her language proficiency level is slightly lower than SHASHA.

The Linguacuisine digital app is, therefore, used as a secondary prompt rather than a primary source of information, which reflects YEDA & SHASHA's advanced understanding of the task, allowing them to rely on prior knowledge and manage their task independently.

Additionally, both learners demonstrate a high degree of autonomy, frequently self-selecting turns. Their interaction is tightly coordinated.

By comparing this extract to prior ones, we observe a change in the locus of epistemic authority: LS 1 does not direct the task flow; rather, it is treated as a resource to be cued, delayed, or muted by the learners. Unlike *Extract 5*, where LS 1's sequence was enforced by one participant (HIL), here it is subordinated to peer coordination.

This form of self-selection is often multi-layered: learners position themselves as next speakers through touchscreen actions (e.g., pausing, pointing), while simultaneously displaying task knowledge or role readiness. Crucially, the app is not prompting them, it is oriented to by learners as part of the turn-taking system. The interface becomes a temporal scaffold learners actively engage with to synchronise participation and manage transition places.

6.1.4. Embodied self-selection: enacting turn initiation through non-verbal means

This section shows how learners accomplish self-selection solely or primarily through embodied actions such as gaze shift, tool use, bodily movement, or handover sequences, often in the absence of verbal cues.

In these cases, self-selection is neither cued by talk nor triggered by the app, but emerges through visible embodied actions that position the learner as next actor. These include initiating the next task step, physically passing an object, or using a gesture to project an upcoming action.

In the following *Extract 7*, two intermediate-level participants, ZHEN (HSK 4) and ALI (HSK 3), engage in step 13 of the during-task phase. ZHEN, who holds a relatively higher epistemic status (K+), conveys this solely through non-verbal actions, such as selecting the required item ('*kuai zi*' - chopsticks) and passing them to ALI.

ω ZHEN   *א* ALI

	ZHEN	ALI
HSK Level	HSK 4 (intermediate)	HSK 4 (intermediate)
Ethnicity	Tajikistani	Tajikistani
Gender	Male	Male
Age Group	18-25	18-25
University Programme	International Relationship (MA)	International Relationship (MA)
Arrived in China	09/2016	09/2016
Length of Stay in China	2.5 years	2.5 years

```
745            ω (0.3) ω

        zhen   ωpresses the

               play button

               on the LS 1ω

746    LS 1    ω*א jie xia lai, qing yong kuai zi kuai su de jiao ban jun yun ω*א

                   next, please use the chopsticks to stir

        zhen   ω gazes at the LS 1--------------------------->>ω

        ali      א gazes at the LS 1--------------------------->>א

747    LS 1    = ω # shi hong yong he yu mi dian fen ω# rong jie kai ω

               =until the brown sugar and corn starch dissolved

→       zhen     ω picks up the chopsticks with left handω

                                             ωunpacks the chopsticks

                                             keeps gazing at LS 1ω

               #fig.28                       #fig.29

        fig
```


28

Extract 7

29

748   ⍵ #(2.0) ⍵

→   zhen   ⍵leans right

              and starts

              stirring in

              the small white

              bowl using

              the chopsticks⍵

      fig   #fig.30

30

749   ⍵ #(1.0) ⍵   א #(1.1) א

→   zhen   ⍵passes the   אtakes over

      ali   chopsticks   the chopsticks

              to **ALI** ⍵   from **ZHEN** א

      fig   #fig.31   #fig.32

31

32

750   (1.1)

      ls 1   puts down

              the bowl,

              finishes

              the non-verbal

              demonstration

751→ **ZHEN**   *kuai zi (.)⍵ # (ran hou) ⍵*

                          . (*continued*).

chopsticks (.) (then)

```
zhen              ⌀ hand gesturing

                  the 'stir' action

                  with RH------>>⌀

fig                  #fig.33
```



```
                                        33                    33a
```

```
752        ℵ*⌀ (13.6)⌀

ali    ℵstarts to

       stir using

       the chopsticks------>>

zhen      ⌀looks at

          the small

          white bowl⌀

753        ⌀ (2.3) ℵ*⌀

ali    ⌀ leans forward,

       RH on LS 1,

       moves to the

       next step ⌀

zhen   >>------->ℵ
```

. (*continued*).

ZHEN and ALI's step 17 begin with ZHEN pressing the play button on LS 1, where the LS 1 provides the instruction in line 746, ZHEN immediately demonstrates his relative more knowledgeable epistemic status (K+) by picking up the chopsticks, unpacking it, and initiating the stirring process from line 747 to 748 (figs. 28, 29, & 30). This non-verbal self-selection action establishes his leadership role in performing the task. Throughout this sequence, ZHEN maintains a steady gaze on LS 1, indicating his orientation toward the technological instructions as a key resource for the task.

Following his initial actions, ZHEN displays consideration for his co-participant, ALI, by passing the chopsticks to him (Fig. 31). This handover is a significant non-verbal gesture, signalling an implicit turn-taking mechanism and providing ALI with an equal opportunity to engage in the task. ZHEN's hand gesture is accompanied by a brief hand motion that mimics the 'stir' action (fig. 33 and a), offering a visual cue to guide ALI's subsequent actions. ALI takes over and completes the stirring process from line 752, demonstrating a seamless transition in roles facilitated through embodied actions rather than verbal communication.

This analysis examines how ZHEN's embodied self-selection signals his intention to take the next action while simultaneously creating space for ALI's participation. In this interaction, non-verbal resources emerge as central to asserting epistemic authority, coordinating task progression, and fostering collaborative engagement in the absence of explicit verbal negotiation. Moreover, the lack of explicit guidance on how to engage with each other or the app compels participants to develop their own strategies for collaboration and task execution. ZHEN's use of non-verbal self-selection not only reflects his understanding of the task at hand but also illustrates how higher-level participants can scaffold their peers' learning experiences through embodied actions.

Additionally, learners may combine non-verbal next-speaker selection with verbal scaffolding to support lower-level participants. *Extract 8* illustrates this interactional phenomenon through a pair consisting of an advanced-level learner, DALI (HSK 5), and an intermediate-level learner, WEI (HSK 4).

Ø DALI ⊆ WEI ⊓ LS 1
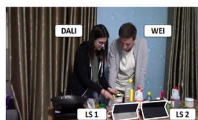
	DALI	WEI
HSK Level	HSK 5 (advanced)	HSK 4 (intermediate)
Ethnicity	Belgian	Ukrainian
Gender	Female	Male
Age Group	18-25	18-25
University Programme	Chinese International Education (MA)	Law (MA)
Arrived in China	09/2017	09/2017
Length of Stay in China	1.5 years	1.5 years

```
055   LS 1    xian zai, qing yong shou ba liang ge ⊓ # suan ban de pi::qu diao ⊓

              now, please use hands peel off two garlic cloves

      ls 1                                    ⊓picks up the garlic---->⊓

      fig                                      #fig.34(LS 1 interface)
```


34
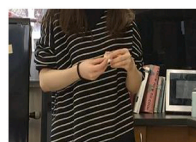
—34a

```
056→          Ø*⊆ #(0.5)  Ø*⊆

      dali    Ø picks up one

              garlic clove

              and gives it

              to WEI, maintains

              gaze to LS 1Ø

      wei       ⊆gazes at the

                LS 1 and takes

                the garlic from

                DALI-------->⊆

      fig         #fig.35
```
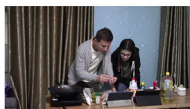
 35   35a

```
057           (0.3)

058→  DALI    pi:: qu diao

              skin:: peel it off

059           Ø*⊆ #(10.0)  Ø*⊆

      dali    Østarts to peel

              off one garlic

              clove------->Ø

      wei       ⊆starts to peel

                the other one

                garlic clove⊆
```

Extract 8

We see the difference between *Extract 8* and *Extract 7* lies in the interaction at line 056, where DALI picks up one garlic clove and hands it to WEI (fig. 35 and 35a). In contrast, WEI takes the garlic clove from DALI but does not proceed; instead, he sustains his gaze on the LS 1 (fig. 35), which potentially suggest his uncertainty about how to proceed and his attempt to follow the non-verbal

demonstration. This pause and WEI's action in line 057 signal to DALI that WEI is struggling, especially considering DALI is an advanced-level participant. As a result, DALI self-selects to repair the situation by repeating the key parts of the cooking instructions with the target vocabulary included – '*suan*' (garlic). Notably, we translate the first instance of '*pi*' in line 058 as '*skin*' rather than '*peel*', since here it functions as a noun, and the following '*qu diao*' is translated as '*peel it off*', to clarify the distinction.

The two participants then proceed with the task, each peeling one garlic clove. Although WEI does not self-select to repeat DALI's repair, he continues the task by peeling the garlic clove alongside DALI. It remains unclear whether WEI is having difficulty understanding the Chinese instructions or if he simply waits for LS 1 to finish its turn. Interestingly, DALI only hands WEI one garlic clove, despite the instruction calling for two, choosing instead to complete the task in a more collaborative manner, which contrasts with the approach taken by LING and LI in *Extract 3*.

The interaction illustrates how a higher-proficiency learner (DALI) can use minimal verbal scaffolding to reorient to a hesitant partner (WEI) and restore momentum in the task. Self-selection here emerges through a multimodal constellation: handover of an object, embodied withholding, gaze monitoring, and a succinct verbal prompt.

In summary, all the four trajectories of CDK learners' self-selections in this section have demonstrate that self-selection was not a one-off act, but part of an emergent participatory trajectory, often evolving from tentative, gesture-based actions to more explicit verbal formulations, marking an expanding semiotic repertoire and increasing orientation to task contingencies.

At the same time, moments of interactional vulnerability—such as hesitations, overlapping self-selections, or mis-coordination—were not uncommon, particularly during the early phases of the task. These instances highlight the emergent and negotiated nature of learner initiative under conditions of limited pedagogical scaffolding. While all groups drew on a shared repertoire of multimodal resources, the fluency and deployment of these varied by pairing, task familiarity, and learner confidence.

In sum, self-selection was not a uniform process but a dynamic, contingent achievement, mediated through talk, gesture, gaze, object manipulation, and digital interface navigation. Even in the absence of a teacher, learners displayed capacity to co-construct the task through embodied alignment, verbal economy, and collaborative timing.

6.2. Integrative commentary: multimodal strategies and interactional consequences of self-selection

Building on the preceding analyses, this section integrates the four analytic trajectories: knowledge-display, sequential-organisation, technology-mediated opportunity, and embodied self-selection, under a unified lens. While each highlights a distinct trajectory of 'how' self-selection is accomplished, the comparison reveals a shared multimodal architecture through which learners coordinate action, claim turns, and sustain participation.

Across all four trajectories, learners mobilised a consistent and interwoven set of strategies to claim turns, direct task activity, and navigate shared epistemic space. These strategies were not only interactionally consequential but also highly multimodal and contextually embedded. The most frequent and recurrent strategies included:

1) Verbal repetition of L2 lexical items after listening to the app audio/video, serving to confirm comprehension and initiate contribution.
2) Reading aloud the Chinese name of the target item from the screen, reinforcing semantic encoding and directing partner attention.
3) Matching visual and textual input—either pairing screen-displayed names or photos with physical objects—thus operationalising task goals.
4) Deictic referencing through gaze, pointing, touching, or moving items (both on-screen and in the physical kitchen space), which functioned as embodied turn-claims and guided joint attention.
5) Confirmation strategies, such as nodding, repeating or echoing correct items, or gesturing toward selected objects, often signalled alignment or closure.
6) Technological interactions, such as pressing or navigating LS1/LS2 buttons, were used not only to control app flow but also to synchronise pair activity.
7) Peer coordination, whether through calling a partner's name, gesturing toward them, or making eye contact, helped manage attention and regulate turn-taking.
8) Embodied movements, including shifting posture, orienting to objects, or transferring tools, were frequently used to assert initiative, align with task sequences, or reallocate roles.

These strategies were often layered—e.g., verbal repetition was accompanied by pointing or gaze, or object handling was paired with a confirmation look—illustrating the multimodal nature of self-selection in the CDK context. Importantly, these strategies were not merely supportive but central to how learners progressed through the task and structured their joint engagement.

First, knowledge-display self-selection shaped epistemic positioning and stance negotiations among learners. By initiating turns to either clarify or assert knowledge, learners actively manage knowledge asymmetries and orient to epistemic roles (K+ or K-). Common strategies included lexical repetition, reading aloud, gaze directing, and object referencing – multimodal configurations that indexed stance and distributed knowledge over time.

Second, sequential-organisation self-selection enables learners to structure task flow and maintain procedural momentum. Learners used verbal instructions (e.g., naming items), object matching (e.g., pairing screen images with real life kitchen items), and embodied behaviour (e.g., touching, moving, handing over, or aligning objects), which served critical to organise task progression. These actions often marked key transitions and prevented interactional drift, aligning participants on shared task ecology. Deictic gestures and task artefacts (e.g., ingredients, icons, written characters) were central to this coordination.

Third, technology-mediated opportunity for self-selection emerged when learners synchronised turns with app-generated prompts. Engagements such as pressing LS 1/LS 2, pausing or replaying clips, or anticipating the app's next move functioned as turn allocation mechanisms. There were often accompanied by verbal or embodied cues, suggesting that the interface acted less as a directive force and more as a manipulable interactional resource.

Finally, embodied self-selection was achieved through gesture, spatial alignment, and artefact handling in the absence of talk. Learners shift posture, passed objects, or performed relevant actions to claim next turns. These embodied cues scaffolded interaction under conditions of linguistic difficulty or high task density, sustaining participation and reducing the need for verbal negotiation.

Together, these self-selection practices shaped the interactional ecology of the CDK. Through verbal and non-verbal means, learners dynamically and collaboratively managed knowledge display, task progression, and role allocation through layered, interdependent resources. Even in the absence of teacher guidance, digital, material, embodied affordances were mobilised to construct coherent, forward-moving joint action.

7. Findings and discussions

This study examines how L2 learners in a Chinese Digital Kitchen (CDK) environment accomplish self-selection in the absence of a teacher or externally structured turn allocation. Using Multimodal Conversation Analysis, we identified four recurrent trajectories of learner-initiated self-selection: 1) knowledge-display self-selection; 2) sequential-organisation self-selection; 3) Technology-mediated opportunity for self-selection, and 4) embodied self-selection. Each trajectory was interactionally situated, sequentially organised, and multimodally achieved, revealing learners' strategic deployment of verbal, embodied, and technological resources to coordinate turns, manage epistemic uncertainties, coordinate task flow, and enact learner agency.

The following sections address how these four trajectories contribute to learners' accomplishment of self-selection (RQ 1) and how the CDK environment mediates these practices (RQ 2).

7.1. Reframing self-selection in the absence of external allocation

Learner self-selection in the CDK environment departs from canonical CA models of floor competition. Instead, turn initiations are embedded in cooperative task management and joint coordination through gaze, gesture, and material or technological resources. This reconceptualises self-selection as a collaborative, trajectory-sensitive process rather than a contest for the conversational flow. In the CDK, turn-taking is not pre-assigned by institutional figures (e.g., teachers), nor does the app itself regulate speaker transitions. Instead, learners negotiate participation through embodied and material cues, creating a distributed ecology of action marked by mutual adaptation and role fluidity. For example, in *Extract 2*, YING self-selects mid-task by pointing to soy sauce while asking '*this one?*', aligning her bodily orientation and speech to self-select without disrupting task flow, showcasing agency through multimodal alignment rather than competitive entry.

Such findings extend prior CA-SLA research (Mori, 2004; Hellermann, 2006) by showing that in technology-mediated, peer-led contexts, self-selection becomes a form of collaborative orchestration, dynamically shaped by the affordances and material environments.

7.2. Knowledge display and the trajectory of epistemic participation

Knowledge-display self-selection foregrounds how learners initiate turns to manage epistemic positioning and mutual under-standing. These self-selections often arise from knowledge asymmetries and are performed through multimodal cues such as verbal repetition, *wh*-interrogatives (Raymond & Heritage, 2013), or deictic gestures.

In *Extract 2*, YING self-selects to question the meaning of *'sheng chou'* (soy sauce) prompted LI's embodied explanation, using the soy sauce bottle and pointing, therefore demonstrating how learners resolve uncertainty and co-construct shared meaning without external guidance. These practices resonate with Waring's (2011) view of self-initiated repair is both an index of learner vulnerability and an opportunity for epistemic alignment. They also reflect Koç's (2023) view of visible cognitive processing in interaction.

Within the CDK, knowledge-display self-selection thus function as both repair and progression, simultaneously addressing lexical uncertainty and propelling the task forward. Learners' epistemic stance-taking not only bridges immediate comprehension gaps but also structures their ongoing interaction trajectories.

7.3. Sequential organisation and the structuring of task flow

Sequential-organisation self-selection captures how learners use turn initiation to structure, maintain, and advance task sequences. Rather than reacting to breakdowns, these actions proactively organise joint activity through verbal proposals, confirmations, and embodied task coordination.

In *Extract 3*, LING self-selects to suggest the next step in the recipe, aligning her talk and gesture with the app's video instruction. The partner's affirmative response confirms alignment, and the pair proceeds with the task. Here, self-selection operates as an interactional mechanism for temporal alignment and joint progression, thereby fostering collaborative rather than unilateral problem-solving dynamics.

This resonates with Hasegawa's (2021) finding that material artefacts (e.g., recipes) and learner positioning shape turn-taking. It also reflects socio-cultural views of learning as mediated action (Vygotsky, 1978; Lantolf & Thorne, 2006), where cognition is

distributed across participants, tools, and tasks. Rather than acting unilaterally, learners in the CDK often co-construct procedural sequences through talk, gesture, and artefact handling, transforming task objects into interactional anchors that stabilise participation.

7.4. Technology-mediated opportunities for self-selection

Technology-mediated opportunities for self-selection refers to how learners orient to the temporal and visual affordances of the Linguacuisine app as part of their turn-construction. Rather than functioning as a passive instructional aid, the app – while non-interactive – functions as a mediator by providing temporal and interactional video and audio prompts. The Linguacuisine app, functions not as a director of action, but as a key interactional resource of both epistemic clarification and independent task management. It does not allocate turns or dictate participation; it structures opportunities for turn initiation through temporal and visual affordances. Learners orient to these affordances as resources for coordination, reinterpreting the app's pacing to sustain joint action. They use its timeline, pause/replay functions, and screen cues to initiate or restructure turns.

In *Extract 6*, YEDA initiates step 1, pauses the app after hearing the word '*qie zi*' (eggplant), and verbally repeat it to confirm comprehension before proceeding. This illustrates how technological features (pauses, replay, skip) become embedded within the turn-taking system, aligning with Lerner (2003) concept of 'environmentally occasioned' turn taking, which is not fixed but rather part of a fluid, negotiated process requiring mutual adjustment among participants and technological resources.

Our findings extend CALL research by showing how technology scaffolds, rather than dictates, learner agency (Jiang & Lai, 2025; Peterson, 2012). The Linguacuisine app provided temporal and visual support that learners actively manipulated to manage participation. The CDK's peer-led, real-world context offers a unique lens to observe learner initiative as a dynamic, multimodal process shaped by both digital prompts and peer actions. Learners' self-selections were neither random nor scripted but affords new temporal frameworks for self-selection, through which learners co-regulate activity and their initiatives.

7.5. Embodied self-selection and material coordination

Embodied self-selection encompasses cases where learners initiate or transfer turns primarily through bodily action, such as tool use, spatial movement, gaze shifts, or gesture, often without talk. These multimodal practices enact participation transitions and facilitate task continuity even under linguistic constraint.

In *Extract 7*, ZHEN moves toward the pot and begins stirring before any verbal exchange, projecting next action and implicitly nominating ALI to continue. ALI's smooth uptake demonstrates shared orientation achieved solely through visible conduct. Similarly, DALI's (*Extract 8*) handover of garlic cloves to WEI, accompanied by a brief repair ('*pi:: qu diao*" – peel it off), show how embodied and minimal verbal cues jointly scaffold less proficient peers.

These practices illustrate what Mondada (2018) calls 'multimodal gestalts' – coordinated configurations of gaze, movement, and material engagement that temporally organise participation. They also reflect Gardner and Mushin's (2017) insight that coordination is often 'seen but unnoticed' until analytically unpacked. The CDK learners' embodied interaction not only complements verbal interaction but can independently facilitate sequential organisation of self-selection for advancing task trajectories and achieving mutual understanding.

By integrating non-verbal self-selection with verbal and digital cues, learner construct a richly multimodal participation framework – underscoring the need for CA-for-SLA research to move beyond linguistic analysis alone (see also Tai & Brandt, 2018).

7.6. Interactional consequences and trajectories

Across all four trajectories, self-selection serve as a dynamic participation mechanism rather than a fixed conversational slot. Knowledge-display and sequential-organisation trajectories tend to exert the most immediate effects on task progression by marking transitions or resolving epistemic ambiguities. Technology-mediated self-selection sustains temporal synchrony with the app, while embodied self-selection provides an alternative, low-resistance entry point for less verbally fluent learners.

Over time, learners demonstrated increasing fluency in combining modalities—moving from tentative verbal queries to seamless verbal-embodied coordination. This suggests a developmental trajectory in how learners manage turn allocation, with practice refining their sensitivity to timing, peer positioning, and technological pacing. Variations across pairs underscore the role of interactional repertoire: some learners relied more on embodied cues, while others foregrounded language or app interaction. These findings mirror Sert's (2017) view of interactional competence as a dynamic, adaptive capacity responsive to local contingencies rather than stable competences.

7.7. Implications and future directions

This study contributes to CA-for-SLA by revealing how learner initiative unfolds in unscripted, multimodal, digital contexts, moving beyond institutional models of turn allocation. It advances our understanding of how self-selection operates in peer-led, digitally mediated environments where no institutional control structures participation. Pedagogically, it supports designing environments that promote flexible, learner-led coordination. For example, incorporating pause/replay functionalities, multimodal strategies, and tasks that encourage spontaneous role negotiation.

Future research should explore longitudinal developments in self-selection strategies, examine cross-linguistic and cross-proficiency differences, and investigate how spatial configurations shape embodied participation. Ultimately, self-selection in the

CDK is not about claiming the conversational floor, but about co-constructing participation through distributed multimodal negotiation among peers, tasks, artefacts, and technological affordances.

CRediT authorship contribution statement

Simin Ren: Writing – review & editing, Writing – original draft, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Paul Seedhouse:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Formal analysis, Conceptualization.

Additional Information

1. The Chinese language has four tones (the first tone is a flat tone, second is a rising tone, third is a dip tone and fourth is a falling tone), but there is no teaching of tones in the CDK task, so this focus on tone learning by the learners is entirely self-nominated and incidental. It is in fact very difficult for L2 learners to pronounce each tone correctly every time, which is the reason why tones are not a part of the task or assessment. **This study did not attempt to focus on tones** for the following reasons: 1) In Chinese/Mandarin, words are either single syllable or are made by putting together two single-syllable words. Each syllable has its own meaning and can have one tone that defines the meaning of the syllable or the word. Tones are used to differentiate words from each other, like consonant and vowel combinations in English. Correct pronunciation regarding tones is therefore important, as is the ability to distinguish tones while listening; 2) However, it can be very challenging for learners whose L1 is non-tonal to not only listen to and distinguish the Chinese tones but also product them precisely; 3) it is also very difficult for us as analysts or researchers to evaluate whether they have achieved the 'native-like' or precise standard of the tones. Therefore, tones remain unmarked in our transcripts; they are only marked when participants decided to learn/repair them.
2. Explanations of LS 1 and LS 2 in Fig. 3 – Research Setting:

It was unusual to have two iPads in this setting, as the Linguacuisine design typically requires only one device with a stable internet connection. However, due to the highly unstable internet connection in the communal kitchens during the data collection, and to prevent participants from waiting excessively for the next instructional video to load, the researcher decided to use two iPads.

One iPad (named **LS1**) was dedicated solely to playing the instructional videos for each cooking step. Since all videos were pre-loaded into the album, this iPad did not require an internet connection and functioned as the primary tool for delivering audio and visual prompts. The second iPad (named **LS 2**) was used for accessing additional resources available on the Linguacuisine app—such as video, audio, and text-based help—whenever participants needed further guidance. A portable Wi-Fi device was placed next to this iPad to ensure a more stable internet connection. (See the two screenshots below for reference).

While the use of two iPads addressed the connectivity challenges, it did introduce certain limitations. For example, if a single iPad had been used without internet connection issues, participants might have been more inclined to engage with the app's additional resources beyond the video instructions. The unreliable Wi-Fi not only affected the ecological design of the digital kitchen (necessitating the use of two iPads instead of one) but may also have shaped participants' interactional choices, as they could have been hesitant to experiment with the app's online features. This does not indicate a flaw in the current study; rather, it reflects how the interactions might have unfolded differently under more stable technological conditions.
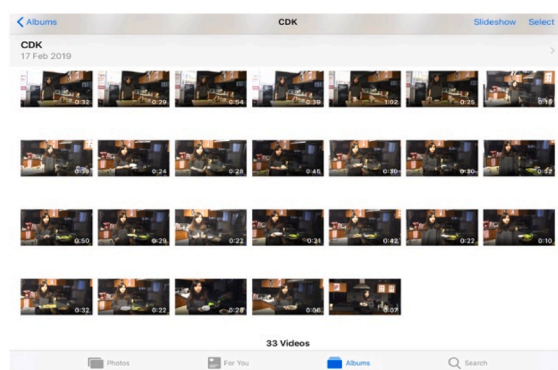
LS 1

LS 2

Appendix A. Transcription Conventions (Jefferson, 2008; Mondada, 2018)

[	Overlap onset
]	Overlap offset
=	Contiguous utterances, or continuation of the same turn by the same speaker even though the turn is separated in the transcripts
(0.2)	The tenths of a second between utterances. Interval between utterances
(.)	A micro-pause (1 tenth of a second or less)
:	Sound extension of a word (more colons demonstrate longer stretches)
–	An abrupt stop in articulation
_	Emphasised in articulation
CAPITALS	Esepcially loud sounds relative to surrounding talk
° °	Talk that is quieter than surrounding talk
> <	Talk that is spoken faster than surrounding talk
< >	Talk that is spoken slower than surrounding talk
↑	Rising intonation
↓	Falling intonation
£ £	Talk uttered in a 'smile' voice
* *	Gestures and descriptions of embodied actions are delimited between two identical symbols (one symbol per participant)
hhh	Audible aspiration
.hh	Audible inhalations
*————≫	The action described continues across subsequent lines until the same
————≫*	symbol (————≫*) is reached
#	The exact moment at which a screen shot has been taken is indicated with # showing its position within turn at talk
fig.	number of the screenshot (from video recordings)

References

Allington, R. L. (1980). Teacher interruption behaviours during primary-grade oral reading. *Journal of Educational Psychology, 72*(3), 371.

Biesta, G. (2020). *Educational research: An unorthodox introduction*. Bloomsbury Publishing.

Brouwer, C. E., & Wagner, J. (2004). Developmental issues in second language conversation. *Journal of Applied Linguistics, 1*(1), 29–47.

Cazden, C. B. (2001). *The language of teaching and learning* (2nd ed.).

Dolce, F. M., & Van Compernolle, R. A. (2020). Topic management and student initiation in an advanced Chinese-as-a-foreign-language classroom. *Classroom Discourse, 11*(1), 80–98.

Eder, D. (1982). *Peer influence on student attentiveness during classroom lessons*.

Ellis, R. (2003). Designing a task-based syllabus. *RELC Journal, 34*(1), 64–81.

Firth, A., & Wagner, J. (1997). On discourse, communication, and (some) fundamental concepts in SLA research. *The Modern Language Journal, 81*(3), 285–300.

Frazier, S. (2007). Conversational structures of 'reports'. *In writing class group work*.

Gardner, R. (2004). Conversation analysis. In A. Davies, & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 262–284).

Gardner, R., & Mushin, I. (2017). Epistemic trajectories in the classroom: How children respond in informing sequences. In *Children's Knowledge-in-Interaction: Studies in conversation analysis* (pp. 13–36).

Hasegawa, A. (2021). Sociomateriality of semiscripted pair-work prompts in beginner-level japanese-as-a-foreign-language classrooms. *The Modern Language Journal, 105*(S1), 65–85.

He, A. W. (2000). The grammatical and interactional organization of teacher's directives: Implications for socialization of Chinese American children. *Linguistics and Education, 11*(2), 119–140.

He, A. W. (2004). CA for SLA: Arguments from the Chinese language classroom. *The Modern Language Journal, 88*(4), 568–582.

Hellermann, J. (2006). Classroom interactive practices for developing L2 literacy: A microethnographic study of two beginning adult learners of English. *Applied Linguistics, 27*(3), 377–404.

Hellermann, J. (2008). *Social actions for classroom language learning*. Multilingual Matters.

Jefferson, G. (2008). Glossary of transcript symbols with an introduction. In G. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13–31). John Benjamins Publishing Company.

Jenks, C. (2013). Are you an ELF? The relevance of ELF as an equitable social category in online intercultural communication. *Language and Intercultural Communication, 13*(1), 95–108.

Jiang, L., & Lai, C. (2025). How did the generative artificial intelligence-assisted digital multimodal composing process facilitate the production of quality digital multimodal compositions: Toward a process-genre integrated model. *Tesol Quarterly*. https://doi.org/10.1002/tesq.3390

Kääntä, L., & Kasper, G. (2018). Clarification requests as a method of pursuing understanding in CLIL physics lectures. *Classroom Discourse, 9*(3), 205–226. https://doi.org/10.1080/19463014.2018.1477608

Kasper, G. (2006). Speech acts in interaction: Towards discursive pragmatics. *Pragmatics and Language Learning, 11*.

Kasper, G. (2009). Categories, context, and comparison in conversation analysis. In *Talk-in-interaction: Multilingual perspectives* (pp. 1–28).

Koç, T. (2023). Requesting clarifications for the meanings of L2 vocabulary items and the emergence of situated cognition and learning. *Language Learning Journal, 52*(4), 380–396. https://doi.org/10.1080/09571736.2023.2179655

Kurhila, S., & Kotilainen, L. (2017). Cooking, interaction and learning: The Finnish digital kitchen as a language learning environment. In P. Seedhouse (Ed.), *Task-based language learning in a real-world digital environment: The european Digital Kitchen* (p. 157).

Kurhila, S., & Kotilainen, L. (2020). Student-initiated language learning sequences in a real-world digital environment. *Linguistics and Education, 56,* Article 100807.

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.

Lauzon, V. F., & Berger, E. (2015). The multimodal organization of speaker selection in classroom interaction. *Linguistics and Education, 31,* 14–29.

Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Ablex Publishing Corporation.

Lerner, G. H. (2003). Selecting next speaker: The context-sensitive operation of a context-free organization. *Language in Society, 32*(2), 177–201.

Long, M. (2015). *Second language acquisition and task-based language teaching*. John Wiley & Sons.

Markee, N. (2000). *Conversation analysis*. Routledge.

Markee, N., & Kasper, G. (2004). Classroom talks: An introduction. *The Modern Language Journal, 88*(4), 491–500.

McHoul, A. (1978). The organization of turns at formal talk in the classroom. *Language in Society, 7*(2), 183–213.

Mehan, H. (1979). 'What time is it, Denise?": Asking known information questions in classroom discourse. *Theory and Practice, 18*(4), 285–294.

Mondada, L. (2018). Multiple temporalities of language and body in interaction: Challenges for transcribing multimodality. *Research on Language and Social Interaction, 51*(1), 85–106. https://doi.org/10.1080/08351813.2018.1413878

Morales, S. (2017). Locating the European digital kitchen. In P. Seedhouse (Ed.), *Task-based language learning in a real-world digital environment: The European digital kitchen* (p. 19).

Mori, J. (2004). Negotiating sequential boundaries and learning opportunities: A case from a Japanese language classroom. *The Modern Language Journal, 88*(4), 536–550.

Mori, J. (2007). Border crossings? Exploring the intersection of second language acquisition, conversation analysis, and foreign language pedagogy. *The Modern Language Journal, 91*, 849–862.

Mori, J., & Zuengler, J. (2008). Conversation analysis and talk-in-interaction in classroomsN. Hornberger (Ed.). *Encyclopedia of Language and Education, 3*, 15–26.

Mortensen, K. (2008). Instructions and participation in the second language classroom. *Doctoral dissertation, Syddansk Universitet)*.

Mortensen, K. (2009). Establishing recipiency in pre-beginning position in the second language classroom. *Discourse Processes, 46*(5), 491–515.

Orletti, F. (1981). Classroom verbal interaction: A conversational analysis. In H. Parret, M. Sbisà, & J. Verschueren (Eds.), *Possibilities and limitations of pragmatics* (pp. 531–549). John Benjamins.

Pavlenko, A., & Lantolf, J. P. (2000). Second language learning as participation and the (re)construction of selves. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 155–177). Oxford University Press.

Peterson, M. (2012). Learner interaction in a massively multiplayer online role playing game (MMORPG): A sociocultural discourse analysis. *ReCALL, 24*(3), 361–380. https://doi.org/10.1017/S0958344012000195

Raymond, G., & Heritage, J. (2013). One question after another: Same-turn repair in the formation of Yes/No type initiating actions. *Conversational repair and human understanding, 30*, 135.

Ren, S., & Seedhouse, P. (2024). Doing language testing: Learner-initiated side sequences in a technology-mediated language learning environment. *Classroom Discourse, 15*(4), 317–352. https://doi.org/10.1080/19463014.2024.2305446https://doi.org/10.1080/19463014.2024.2305446.

Rodriguez, & Wilstermann. (2018). Learner initiative in the Spanish as a foreign language classroom: Implications for the interactional development. *Hacettepe University Journal of Education, 33*, 113–133.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*(4), 696–735.

Sahlström, F. (1999). Up the hill backwards: On interactional constraints and affordances for equity-constitution in the classrooms of the Swedish comprehensive school. *Doctoral dissertation, Acta Universitatis Upsaliensis)*.

Schegloff, E. A. (1987). Analyzing single episodes of interaction: An exercise in conversation analysis. *Social Psychology Quarterly*, 101–114.

Schwab, G. (2011). From dialogue to multilogue: A different view on participation in the English foreign-language classroom. *Classroom Discourse, 2*(1), 3–19.

Seedhouse, P. (2009). The interactional architecture of the language classroom. *Bellaterra Journal of Teaching and Learning Language and Literature, 1*(1), 1–13.

Seedhouse, P. (Ed.). (2017). *Task-based language learning in a real-world digital environment: The European Digital Kitchen, 4*. Bloomsbury Publishing.

Seedhouse, P., Heslop, P., Kharrufa, A., Ren, S., & Nguyen, T. (2019). The linguacuisine project: A cooking-based language learning application. *The EuroCALL Review, 27*(2), 75–97.

Sert, O. (2017). Creating opportunities for L2 learning in a prediction activity. *System, 70*, 14–25.

Sidnell, J. (2010). *Conversation analysis: An introduction*. John Wiley & Sons.

Sinclair, J., & Coulthard, M. (2013). Towards an analysis of discourse. In *Advances in spoken discourse analysis* (pp. 1–34). Routledge.

Tai, K. W. H., & Brandt, A. (2018). Creating an imaginary context: Teacher's use of embodied enactments in addressing learner initiatives in a beginner-level adult ESOL classroom. *Classroom Discourse, 9*(3), 244–266. https://doi.org/10.1080/19463014.2018.1496345

Ten Have, P. (2007). *Doing conversation analysis*.

Van Balen, J., Gosen, M. N., de Vries, S., & Koole, T. (2024). Peer-to-peer talk in whole-classroom discussions. *International Journal of Educational Research, 125*, Article 102354.

Van Lier, L. (2014). *Interaction in the language curriculum: Awareness, autonomy and authenticity*. Routledge.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Walsh, S. (2002). Construction or obstruction: Teacher talk and learner involvement in the EFL classroom. *Language Teaching Research, 6*(1), 3–23.

Walsh, S. (2006). *Investigating classroom discourse*. Routledge.

Waring, H. Z. (2008). Using explicit positive assessment in the language classroom: IRF, feedback, and learning opportunities. *The Modern Language Journal, 92*(4), 577–594.

Waring, H. Z. (2009). Moving out of IRF (Initiation-Response-Feedback): A single case analysis. *Language Learning, 59*(4), 796–824.

Waring, H. Z. (2011). Learner initiatives and learning opportunities in the language classroom. *Classroom Discourse, 2*(2), 201–218.

Wells, G. (1993). Reevaluating the IRF sequence: A proposal for the articulation of theories of activity and discourse for the analysis of teaching and learning in the classroom. *Linguistics and Education, 5*(1), 1–37.

Young, R. F., & Miller, E. R. (2004). Learning as changing participation: Discourse roles in ESL writing conferences. *The Modern Language Journal, 88*(4), 519–535.