

Chinese Vocational College EFL Students’ Multimodal Resource Use for English Dialogue Learning

Lan Wang

Student of Doctoral Degree in Linguistics Program, Suan Sunandha Rajabhat University

E-mail: S68584952002@ssru.ac.th

Suwaree Yordchim

Department of Linguistics, Graduate School, Suan Sunandha Rajabhat University

E-mail: suwaree.yo@ssru.ac.th

Suphat Sukamolson

Department of Linguistics, Graduate School, Suan Sunandha Rajabhat University

E-mail: ssuphatz@gmail.com

Sarada Jarupan

Department of Linguistics, Graduate School, Suan Sunandha Rajabhat University

E-mail: Sarada.ja@ssru.ac.th

Kirk Roger Person

Department of Linguistics, Graduate School, Suan Sunandha Rajabhat University

E-mail: Kirkroger.pe@ssru.ac.th

Abstract

Oral English is still a major weakness for Chinese vocational college students, the article explores the use of multimodal resources by Chinese vocational college students for English dialogue learning from the perspective of multimodal conversation analysis. 81 non-English major freshmen participated in the study as research participants. Using multimodal qualitative analysis a two-month data collection was conducted. The results show that students in order to be able to read and recite English dialogues, will identify sentences and guide with a finger or a pen, scan the QR code audio on the textbook with a cellphone, listen and imitate the dialogue along and gaze at partner’s eyes and interact with the partners. Research has shown that Chinese vocational college students unconsciously utilize their own multimodal resources in the process of English dialogue learning. The use of these multimodal resources can effectively assist students in learning and consolidating English dialogues, providing practical references for the oral teaching reforms of English or other languages.

Keywords: multimodal resource use, English dialogue, English learning

Introduction

In China, writing, grammar, and translation are frequently given more weight in traditional English teaching and assessment than English speaking (Zhang & Head, 2010; Butler et al., 2020). Consequently, Chinese college students’ spoken English proficiency is still a serious shortcoming, especially vocational college students. English teaching at Chinese vocational colleges emphasizes that English serves professional skills and career scenarios. English dialogue learning should pay more attention to its applicability and practicality. The process of students practicing English dialogues on their own is a process of repeatedly reading, understanding, memorizing, and communicating unfamiliar language materials into intrinsically comprehensible input. (Krashen, 1985). This process is subtle and imperceptible. Teachers can use students' multimodal learning habits to change their teaching



methods. The research will be exploring the following research questions within the framework of sociocultural theory. (Lantolf & Beckett, 2009) Will Chinese vocational college students unconsciously use multimodal resources to assist learning in the process of English dialogue learning? How do they use multimodal resources to assist in learning English dialogues from the textbook?

Research Objectives

1. To observe whether Chinese vocational college students use multimodal resources in the process of English dialogue learning.
2. To explore the features Chinese vocational college students use multimodal resources in the process of English dialogue learning.

Scope of the Research

Population and Time Scope

The research involved 81 first-year, non-English major students from a vocational college in Hunan, China. The video data collection was conducted between November and December 2025 in China.

Literature Review

As a core analytical framework for second language acquisition, sociocultural theory criticizes the purely cognitive approach of viewing language acquisition as a paradigm of internal processing within individuals. (Lantolf & Beckett, 2009) The successful acquisition of language is entirely born from social interaction behaviors such as classroom and peer negotiation and then internalized into individual language abilities. Language is just one type of cultural intermediary tool, and all ideographic symbols such as gestures, eye contact, posture, and multimedia serve as intermediaries.

The existing research on multimodal language acquisition within the framework of sociocultural theory focuses on English majors and undergraduate comprehensive university students. (Lin, 2017; Siegel & Seedhouse, 2019; Witte, 2023) There is a scarcity of micro-classroom empirical evidence targeting non-English major groups in vocational colleges with weak foundations and significant oral difficulties, making it difficult to adapt to the teaching positioning of vocational English service industry scenarios and emphasizing practical dialogue. In the past, most studies only analyzed physical modalities such as gestures and eye contact, ignoring localized digital intermediary tools such as QR codes and mobile audio in textbooks and failing to incorporate digital media into the “cultural symbol intermediary” system of social and cultural theory for analysis. Rodríguez and Evnitskaya (2024) have focused on teachers’ multimodal scaffolding behavior, lacking systematic observation of students’ spontaneous and unconscious use of multimodal resources for peer negotiation dialogue, and failing to reveal the complete social behavior chain of learners’ autonomous use of multimodal symbols to construct meanings.

Research Methodology

Due to multiple factors having an impact on college students’ English-speaking ability (Chen et al., 2025), the process of dialogue learning is the co-construction of social action of their English-speaking interaction. (Lantolf & Beckett, 2009) This study adopts multimodal qualitative analysis to observe the speakers’ conversational interaction.



Population and Sample

The population of this research was 132 non-English major freshmen from a vocational college in Hunan, China. A total of 81 participants were willing to finish the oral conversational practice tasks. Of these 81 participants, 45 used multimodal resources during the English conversational practice. All participants were informed and gave consent prior to data collection and used pseudonyms (such as Student J) and vague photos to protect their privacy. The research data includes 15 hours of classroom observation, but as the focus of the study is on the process of students using multimodal resources to learn English dialogues, the study only obtained valid video clips of students using multimodal resources to learn English dialogues.

Data Collection Method

This study collected video data of English classroom students' English dialogues and learning processes through video. The video includes students' multimodal resources; students practice independently and practice or negotiate with their partners in a dialogue learning process. The video clips of nonverbal symbols (such as gaze, finger or pen navigation, and body posture) that integrate language expression during the process of dialogue learning are effective data for this study.

Data Analysis

Utilizing multimodal conversation analysis (MCA) and transcription conventions (TC) (Hellermann, 2008) as tools to transcribe and screenshot conversation videos of observing students' multimodal movements.

Research Results

Preliminary analysis of the data corpus shows that not all students use multimodal resources to learn English dialogues. A total of 49 occurrences were identified from 15 hours of first-hand data. Three typical dataset extracts have been chosen for this investigation.

Identifying Sentences and Guiding with a Finger or Pen

Excerpt 1



Figure 1: J uses a finger to identify the sentences
(video-2025-1110/0:07:54)



Figure 2: H uses a pen to identify the sentences
(video-2025-1110/0:00:07)

01Student J: I could never be so successful without your help ((Point with the finger at the English dialogue in the textbook))

02Student K: I°Oh::: (...) ((Stare at the textbook))

03Student J: I am so sorry for being late again ((Point with the finger at the English dialogue in the textbook))



04Student K: Er:::(...) ((Stare at the textbook))

05Student J: Please forgiving me one more time ((Point with the finger at the English dialogue in the textbook))

06Student K: Oh::: Never mind °°((Stare at the textbook))

07Student J: Sorry for that ((Point with the finger at the English dialogue in the textbook))

Figure 1 shows that 2 students are having dialogue interaction; due to unfamiliarity with the dialogue, they are both staring at the textbook without face-to-face interaction. Student J uses a multimodal resource (a finger) to point to the dialogue sentence to avoid misreading (01). Student K replies to him in a quiet voice (02). Student J constantly points at the textbook with a finger (03)(05)(07). Based on dialogue transcription, we do not observe pause, sound stretch, and code-switching. It can be concluded that Student J’s speech accuracy, fluency, and completeness are very high, while Student K usually speaks in a low voice and stretches the sounds (04) (06), indicating his unfamiliarity and hesitation. It is worth mentioning that at turns 01, 03, 05, and 07, Student J kept pointing his finger at the dialogue. After class, when Student J was asked why he would use a finger to point at the dialogue in the textbook, he said that it not only prevented misreading but also helped to concentrate and avoid distraction. Figure 2 demonstrates that another student H (in dark coat on the left) showed the action of recognizing sentences with a pen, negotiating with his seatmate.

Scanning the QR Code Audio for Listening and Imitating

At present, English textbooks in Chinese vocational colleges are equipped with audio QR codes. When students encounter dialogues that they cannot read and speak, they can scan the code with their cellphones to listen to and imitate the audio of the conversation.

Excerpt 2



Figure 3: *L uses a cellphone to scan the QR code*
(video-2025-1129/0:05:49)



Figure 4: *L uses a cellphone to listen and imitate the dialogue*
(video-2025-1129/0:05:53)

Excerpt 2 presents a student (Student L) is utilizing a cellphone to scan the QR code on the textbook (Fig. 3). Then, she moves the cellphone near her left ear to listen to and imitate the pronunciation and intonation of the dialogue (Fig. 4). These multimodal resources, including textbooks, audio, QR codes, and cellphones, create multiple channels for students to learn English dialogues. Student L does not need the help of her partner or her English teacher from the pictures we can find that.

Gazing at Partner’s Eyes and Interacting with the Partners

Students utilize a finger or a pen to identify sentences to avoid misreading and distraction in the process of dialogue learning we have mentioned (Excerpt 1). Students scan the QR code and read the sentences from the textbook (Excerpt 2), while Excerpt 3 shows students’ gradually proficient conversational reading.

Excerpt 3



Figure 5: M reads the dialogue
(video-2025-1127/0:00:28)



Figure 6: M closes the textbook and
interacts with a partner
(video-2025-1127/0:00:30)



Figure 7: O gazes at M and interacts with M
(video-2025-1127/0:00:38)

- 01 Student M: Ok here is my (...) thank you for that
02 Student M: ((Point with her finger at the sentence on the textbook, Gaze at her partner and smile))(0.3)
((close the textbook))
03 Student M: Hi::: long time no see(.)my::(...)
04 Student O: Hello (0.3) nice to see you again how are you^{oo}((gaze at studentM))
05 Student M: Very well thank you(.) how is your (...)
06 Student O: It is going Er: Er: fine

The beginning of the sequence is initiated by Student M (01). At the 2nd turn of the sequence, Student M points with her finger at the sentence on the textbook (Fig.5), gazes at her partner with a smile, and pauses less than 0.3 seconds. Then she closes the book (Fig.6) and interacts with Student O (02). Different from Excerpt 1 and Excerpt 2, students gaze at their textbook and communicate with their partners. Student M stared at her partner, Student O, directly rather than at the textbook. The multimodal resource (closing the textbook and gazing at her partner) indicates that she is familiar with the dialogue and ready to interact. Student M initiates the turn (03), greeting Student O. There are 3 adjacency pairs emerged at 03-04, 04-05, and 05-06 that construct the process of dialogue learning. Student O gazes at Student M at the 4th turn (Fig.7). From the series of multimodal actions, it can be inferred



that Student M and O has transformed from practicing dialogue to simulating dialogue through role-playing, that is, they have naturally transformed the comprehensible input of English dialogue learning into dialogue output through the use of multimodal resources.

Discussion

Chinese vocational college students use multimodal resources in the process of English dialogue learning. Most of the students didn't realize which kinds of multimodal resources they used but concentrated on whether they had mastered the content of the English dialogue. Lantolf and Beckett (2009) argue that the essence of learning is to participate in social activities and utilize cultural symbol resources, rather than simply “receiving language knowledge.” Scaffolding is a unique social negotiation behavior in the classroom. And the features of Chinese vocational college students using multimodal resources in the process of English dialogue learning, the successful acquisition of language is entirely born from social interaction behaviors such as classroom peer negotiation and then internalized into individual language abilities. (Lantolf & Beckett, 2009) The outcome of the study mainly reflected in the following aspects: Guiding with a finger or a pen to identify sentences; Scanning the QR code audio for listening and imitating; Gazing at the partner's eyes and interacting with the partners. EFL classroom conversation design should not only focus on students' oral English output but also actively build multimodal interactive scenarios: encouraging gestures, visual teaching aids, group physical cooperation dialogue, and consciously using multimodal scaffolds to activate peer assistance. We accept multimodal semantic behaviors observed in student conversations as a normal social-developmental process of language acquisition, rather than a sign of expressive deficiency.

Conclusion

This study confirms that these body postures can serve as cognitive scaffolds, helping students transition from language input (reading and listening repetition) to language output and negotiative meaning-making (closing textbooks and making eye contact). However, the coverage of the sample was only collected in local areas, and the duration of data collection in this study is limited, the place scope and duration time can be expanded in the future. The entire process of dialogue learning is the application of their own multimodal resources and the process of peer conversation interaction, meaning negotiation and language co-construction. EFL students rely on others' multimodal symbol support to complete English conversations beyond their independent abilities within their own Zone of Proximal Development. After long-term repetition of this social behavior, the external multimodal framework gradually internalizes, and students will spontaneously use gestures, pauses, stress, and other modalities to improve their expression when communicating independently. Teachers can use students' multimodal learning habits to change their teaching methods. The multimodal resources of this study involve gaze, gestures, and digital tools. Language learning is a practical behavior that combines sociality and embodiment, integrating language expression, nonverbal symbols (gaze, finger, or pen pointing), and physical carriers (textbooks, mobile phones).



References

- Batlle Rodríguez, J., & Evnitskaya, N. (2024). Teachers’ multimodal resources for delegated peer repair: Maximizing interactional space in whole-class interaction in the foreign language classroom. *The Modern Language Journal*, 108(1), 297–321. <https://doi.org/10.1111/modl.12910>
- Butler, Y. G., Lee, J., & Peng, X. (2020). Failed policy attempts for measuring English speaking abilities in college entrance exams: Cases from China, Japan, and South Korea. *English Today*, 38(4), 271–277. <https://doi.org/10.1017/s0266078420000346>
- Chen, L., Yordchim, S., & Sukamolson, S. (2025). The use of mobile app (Fluency in English) to enhance the speaking English ability among Chinese college students. *Edelweiss Applied Science and Technology*, 9(3), 726–733. <https://doi.org/10.55214/25768484.v9i3.5286>
- Hellermann, J. (2008). *Social actions for classroom language learning*. Multilingual Matters. <https://doi.org/10.21832/9781847690272>
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.
- Lantolf, J. P., & Beckett, T. G. (2009). Sociocultural theory and second language acquisition. *Language Teaching*, 42(4), 459–475. <https://doi.org/10.1017/S0261444809990048>
- Lin, Y. L. (2017). Co-occurrence of speech and gestures: A multimodal corpus linguistic approach to intercultural interaction. *Journal of Pragmatics*, 117, 155–167.
- Siegel, A., & Seedhouse, P. (2019). Conversation analysis and classroom interaction. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 2–8). John Wiley & Sons. <https://doi.org/10.1002/9781405198431>
- Witte, A. (2023). Bringing the body into play: The corporeal aspect in second language acquisition. *The Modern Language Journal*, 107(3), 693–712. <https://doi.org/10.1111/modl.12861>
- Zhang, X., & Head, K. (2010). Dealing with learner reticence in the speaking class. *ELT Journal*, 64(1), 1–9. <https://doi.org/10.1093/elt/ccp018>

