

Chapter 4

Research Methodology



Abstract The research methodology in L2 pragmatics has long been a source of concern, and with legitimate reason: how data is acquired and evaluated is critical in deciding what conclusions may be derived (Taguchi & Roever, 2017). Furthermore, methodologies must be chosen in order to best answer the study questions (Kasper, 2008). As a result, this chapter focuses on the research methods used in the current study. It is further divided into six main sections: (1) descriptions of the participants; (2) instrumentation; (3) data collection; (4) data analysis; (5) inter-rater reliability; and (6) ethical issues. The following details will be displayed.

Keywords Participants · Instrumentation · Data collection · Data analysis · Inter-rater reliability · Ethical issues

4.1 Participants

The participants were 143 Chinese students of English, divided into three groups (see Table 4.1). The 51 test-takers in Group 1 (G1) were all third-year English major undergraduates studying at some universities in China; they had a relatively low English proficiency (EP) level as they had only passed the Test for English Majors, Band 4 (TEM-4); these students had never lived or studied in a country where English was spoken. Group 2 (G2) included 59 Chinese master students with English majors (but with no study-abroad experience; they had passed the TEM-8, China's highest national English test. Group 3 (G3) consisted of 33 Chinese master and doctoral degree students of Chinese language and literature, world history, philosophy, accounting, management, business, and educational psychology; at the time of data collection, they were enrolled in various study-abroad programs and had previously taken either the TOEFL or the IELTS. Given that the TEM-4 is easier than the TEM-8, the proficiency levels of G2 and G3 were considered advanced, while the G1 level was considered intermediate.

At the start of the experiment, a background information survey was administered to the participants, which included age, gender, contact information, length of time and motivation during English learning, proficiency scores, and length of abroad studying or living experience. Several individuals were excluded from the

Table 4.1 Participant information

	Group 1 (n = 51)	Group 2 (n = 59)	Group 3 (n = 33)
Average age (range)	21.08 (20–23)	23.32 (23–25)	27.50 (22–36)
Gender (male: female)	4:47	7:52	9:24
Length of studying English (SD)	12.18 years (2.33)	14.00 years (2.59)	15.33 years (6.65)
Proficiency levels	TEM 4 Average: 65.50 (SD = 5.56; range: 60–80)	TEM 8 Average: 69.00 (SD = 5.30; range: 60–82)	TOEFL (n = 6) Average: 89.33 (SD = 3.77; range: 83–94); IELTS (n = 27) Average: 6.83 (SD = 0.24; range: 6–8)
Length of study-abroad experience	None	None	Average: 10.60 months (SD = 7.43, range: 2–27 months)

final data set because they did not complete one or more of the testing tasks. All of the experimental tasks were completed with a 95% completion rate.

A pilot study was carried out with 41 native American speakers (30 females and 11 males, average age: 23.49, SD = 4.64), whose responses were synthesized as baseline data of target-like responses throughout each routine task.

4.2 Instrumentation

The present study employs a mixed-method, stimulus-led approach using internet-based animated movie sites (www.nawmal.com). During the testing stages, the animated scenarios are created in tandem with elicitation tasks to assess learners' routine performance across various tasks. This animation task can provide more prompts (i.e., animation of interactive context, movement, and images of interlocutors) that increase the degree of naturalness (Félix-Brasdefer, 2010; Ren, 2015), and share the cheerful practicality of DCTs, which is primarily reflected in the ease of administration for gathering large amounts of comparable data under controlled conditions from a large number of respondents in a relatively short period of time. It is widely acknowledged that “natural data is often held up as the ‘gold standard’ of L2 pragmatics data” (Taguchi & Roever, 2017: 119). This task is intended to provide standardized computer-animated, audio-visual input to all participants, guaranteeing the comparability of learners' performance in routine output under differential grouping. Furthermore, it is generally established that DCT elicits offline knowledge (Félix-Brasdefer, 2010), which cannot substitute actual language use in real-world communication. Instead, because its major focus was not on participants'

real-life pragmatic use, but rather on their offline pragmatic competence with regard to target-like routine output that well fits diverse actual situational contexts, this computer-animated activity effectively remedied this deficiency.

4.2.1 Computer-Animated Production Task

To assess learners' abilities to produce routines, the Computer Animated Production Task (CAProT) was employed. The experimental situations were entirely based on Bardovi-Harlig (2009)'s research that targeted expressions for which learners demonstrated low production. The "stimulus-led oral" (Halenko, 2018: 146) CAProT consisted of 13 and 19 target routine scenarios for initiating and responding to utterances, respectively (see Appendix 2 and 3). Figures 4.1 and 4.2, for example, show two immobile screenshots of CAProT scenarios created through this animation technology. The scenarios included a number of animated actual situational settings as well as an initiating utterance by the "American speaker", to which the "Chinese hearer" had to respond by "engaging in a brief, single-turn interaction with the animated higher-status characters" (Halenko, 2018: 146). The characters portrayed individuals the students can encounter in their daily life, such as an academic tutor, a teacher or classmate on a university campus, a salesclerk at a clothes store, and so on.

To be more explicit, after an introductory instructional slide, all participants were invited to first observe (and read) the background of each scenario, which had been converted into a short movie. After a 5-s delay, the animated interlocutor would appear and instruct learners to initiate a conversation or respond to the speakers' utterances. Following that, participants were expected to offer an oral response in the form of either initiating responses or responding to utterances, as directed by the contextual reminders. The learners were then given a 30-s timed interval (20 s for replying and 10 s for the gap between two items) to respond before the scenario was automatically shown.

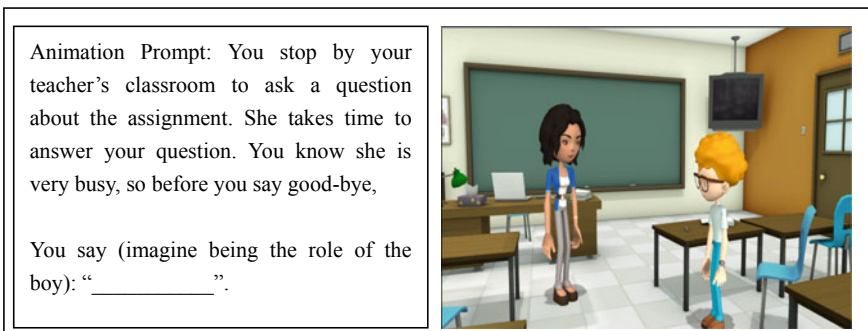


Fig. 4.1 An example scenario for initiating utterances

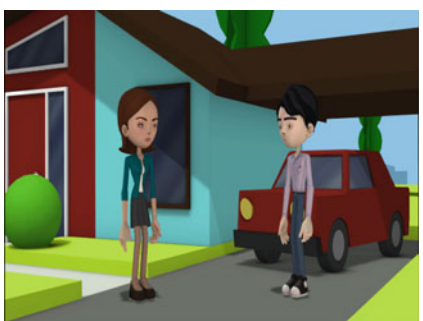
<p>Prompt: The boy gives the girl, his classmate, a ride home. He lives in the building next to hers.</p> <p>Animation: She gets out of the car and says, “Thanks for the ride.”</p> <p>You (imagine being the role of the boy) say: “_____”.</p>	
---	--

Fig. 4.2 An example scenario for responding to utterances

In terms of the initiating task, participants must actively start a conversation after reviewing contextual information and topic requirements in order to fulfil the situational information requirements. As shown in Fig. 4.1, the Chinese youngster (on the right, you) needed to express his gratitude to his American teacher (on the left) for taking up her time to answer his questions before he was ready to leave the classroom based on situational information.

The responding task will display the initiated utterances of the speaker (the American girl on the left), and the listener (“you”, a Chinese boy on the right), should not only respond to the girl’s gratitude but also meet the requirements of contextual information, such as “you two are classmates who live nearby”.

4.2.2 *Computer Animated Recognition Task*

The recognition task is a multiple-choice DCT in which participants must choose the most situationally appropriate response from four choices (i.e., Roever, 2005, 2012). Using animation technology, learners were primarily needed to complete the “visual-audio” computer-animated recognition task, in which the prompts were initially shown in a short movie, with written captions appearing at the bottom of the screen. Following an introductory instructional slide, learners were instructed to observe and view each scenario’s background and options on the left, and they were then needed to choose the most appropriate option from four selections online. Following a 10-s pause, the animated interlocutor would be awakened and begin a fresh dialogue, with another 10-s pause for recognition to complete (Fig. 4.3).


<p>Prompt: Carrie (left) has done some shopping at a grocery store.</p> <p>Animation: The salesclerk (right) at the cash register has just finished packing her groceries and gives her the bags.</p> <p>What would the salesclerk probably say?</p> <p>A. ‘All yours.’</p> <p>B. ‘There they are.’</p> <p>C. ‘Here you go.’</p> <p>D. ‘Please.’</p>	
--	--

Fig. 4.3 An example item for routine recognition: "Shopping grocery"

4.2.3 Computer Animated Comprehension Task

This computer-animated comprehension task (CACT) was also delivered via an animated movie website mentioned above. Several scenarios were produced in order to weaken “the potential for learners to infer meaning from contexts provided by test stimuli” (Bardovi-Harlig, 2014: 43). The prospective target expressions were selected from prior L2 pragmatics investigations (Bardovi-Harlig, 2014; Roever, 2005, 2012), and included *Here you go*, *All yours*, *That works for me*, *For here or to go*, *Do you think you can make it*, *Excuse the mess*, and *Thanks for having me*, on which learners demonstrated both low production and recognition. All of the expressions had a nontransparent compositional meaning and were difficult to identify and produce. Figure 4.4 shows a still snapshot of one of the testing scenarios created with this technique.

To match the modes over the whole task, each targeted expression was shown both aurally and visually twice with a 0.5-s timed interval. Following an initial instructional slide, all respondents were instructed to deliver an oral answer from four alternatives while seated in one-row intervals to prevent the disruption of overlapping noises. All participants had 30 s to finish each task and a 10-s timed period to react before the next scenario was presented automatically. This approach was originally illustrated using a practice animation scenario prior to the formal test phase. All of their oral replies were videotaped by the computer terminal equipment.

4.2.4 Computer Animation Perception Task

Based on the preceding tasks, this computer-animated perception task chose five pairs of routines that were similar in form, meaning, or function but not identical. The boy (on the right) will ask the girl (you, the respondent) a series of questions, such as


<p>Instructions: “Choose ONE answer that best describes your context knowledge. ‘All yours.’ (.5s) ‘All yours.’”</p> <p>A. I don’t know having heard this expression before.</p> <p>B. I have heard this expression before, but I don’t know what it means.</p> <p>C. I have heard this expression before, and perhaps it means _____.</p> <p>D. I know this expression. It means__ and use it to give an example in a concrete actual situational context.</p>	
---	--

Fig. 4.4 An example item for routine comprehension: “All yours”

“Do you believe *Nice to meet you* and *Nice to see you* can be used interchangeably? If this is the case, only say Yes. If not, please describe the various contexts in which these two phrases can be used”. The girl (“you”) should reply “yes” or “no” for the first time; if the answer is “yes”, the problem directly ends; if the answer is “no”, you must demonstrate the precise distinctions between the two routines about their functional applicability in the particular actual situational context. The respondents were then given a 20-s time limit to answer the question before the next paired routines presented automatically on the screen (Fig. 4.5).


<p>The boy asks, “Do you think ‘Nice to meet you.’ and ‘Nice to see you.’ can be used in the same context? If so, only answer ‘Yes’. If not, please point the different situations these two expressions can be applied to respectively.”</p> <p>You (imagine the role of the girl) say, “_____.”</p>	
---	--

Fig. 4.5 An example item for routine perception: “Nice to meet/see you”

4.2.5 Computer Animated Retrospective Review

The utilization of multiple data sources in this investigation follows a trend observed in recent studies; that is, a series of computer-animated elicitation tasks were initially used to elicit learners' pragmatic knowledge, and then a follow-up retrospective interview was conducted to gain insights into learners' responses. The animation was also used in this retrospective interview, as seen in Fig. 4.6. The respondent envisioned himself/herself as the youngster on the left, delivering his/her replies as soon as they heard the questions displayed on the screen.

This test was constructed with the following levels in mind to elicit data from researching the Chinese EFL learners' (without abroad residence) cognitive processes engaged in their routine performances: (1) learners' attention across all tasks; (2) task difficulties; (3) L1/L2 preference to assess the degree of L1-driven transfer; and (4) the major source of prior context knowledge that may be controlled for routine completion. When administering this interview, participants were allowed to offer their replies in Chinese, capable of interacting their views more clearly.

4.3 Data Collection Procedure

By the end of the summer semester in 2019, the entire research has been completed properly. The collection procedure began in May 2019 and ended in November 2019, at which time the whole process was divided into two sections nationally and internationally. Prior to the experiment, each participant was asked to provide informed consent to the collection of oral data for research purposes. They were also informed that the holistic research project would only be exploited for scientific research, their personal information would not be shared, and that their oral responses would be kept confidential.

After agreeing to participate in the experiment, each participant was requested to complete a personal background questionnaire (see Appendix 1 for details). Before

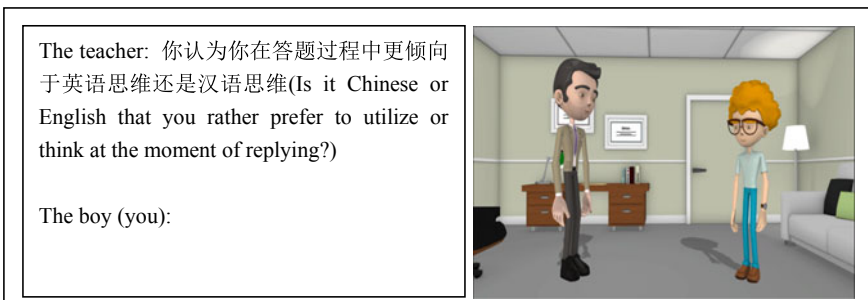


Fig. 4.6 An example for retrospective interview: “L1/L2 Preference”

the formal start of each test phase, the researcher will explain the relevant test requirements. Because the proficiency level of learners at home and abroad is quite high, the researcher's short rundown of pragmatic routines will assist participants in clarifying the assessment goals. It should be reminded that the example questions produced ahead of time are just designed to demonstrate the test objective; learners' replies will not be recorded or scored.

The experimental site is mostly done in five Chinese universities at home, whereas data for Chinese EFL learners overseas is primarily collected during the researcher's visit to the United States. All potential study-abroad participants were distributed and collected via the online questionnaire website due to technical challenges with the experimental location and requisite equipment.

A survey of 110 at-home EFL students was conducted as a consequence of the follow-up retrospective research, which mostly focused on improving the pragmatic skill of Chinese English learners with no prior overseas experience. Overall, the cross-sectional approach took roughly 50–60 min for all participants to engage in the overall phases of experiment conduction.

4.4 Data Analysis

4.4.1 Coding for Routine Production

In this task, distinguishing different aspects of pragmatic knowledge is highly recommended, in contrast to the overall rating. The evaluation system, adapted from Bardovi-Harlig (2019), allows us to fully understand the impact of proficiency and study-abroad experience on learners' production of routines. As a result, for participants' routine manifestations, two mastery levels of actual situational and prior context) knowledge were assessed separately.

Routines are frequently linked to contexts and speech acts, which are “two basic pragmatic constructs” (Bardovi-Harlig, 2019: 47). In such cases, learners' mastery of actual situational context knowledge was assessed based on their comprehension of contextual information and the consistency of the target speech act with a felicitous pragmatic strategy. To assess the learners' mastery of prior context knowledge, a seven-point rating scale was used, with scores ranging from zero (inconsistent or no response) to three (perfectly appropriate), as displayed in Table 4.2 (identical to statements in Wang, 2022).

An example of a learner's initiating utterance was used to demonstrate the coding criteria for routine production more clearly. Above all, “*Excuse me, do you have a time?*”, said by the respondent, met all of the requirements of the actual situational context and thus received 3 points, because it can be inferred at least in these aspects: (1) this learner did indeed interpret contextual information, (2) the target request speech act “request” with proper pragmatic strategy was precisely employed. In the area of prior context knowledge, an uncountable noun *time* was used in place of the

Table 4.2 Rating band for routine production

Level	Rating criteria	Score
Mastery of ASC knowledge	Fully aware of key information provided in the video and implemented the respondent's role (if not, the score is 0)	1'
	Is the speech act consistent? (if not, the score is 0)	1'
	If stated, is the same pragmatic strategy used?	1'
Mastery of PC knowledge	The same content as NSs, almost perfectly appropriate	3'
	Alternative appropriate wording, or slightly non-native-like with verbosity	2.5'
	Routines that do not mention core content but are acceptable	2'
	Attempted lexical core, slightly non-native-like for minor interlanguage grammatical errors	1.5'
	Non-native like for strange wording, major grammatical errors	1'
	Utterances that are incomplete but can be judged	0.5'
	Impossible to understand, totally inconsistent with NSs' responses, or no response	0'

Note ASC, actual situational context; PC, prior context

countable noun *minute* in this response, thus scoring 1.5 points in this section and 4.5 points overall.

4.4.2 Coding for Routine Comprehension

Learners' routines were evaluated based on two aspects in the same task, namely, meaning and use: explicitly stating the definition of a particular routine expression based on prior knowledge and specifying its usage in a concrete actual situational context. Learners' definitions, derived from their prior knowledge, were assessed and coded as "plausible", "implausible", and "no recognition". Plausible definitions comprised all the meanings listed by 41 native speakers. Implausible definitions included *It's up to you* for *All yours* and *To stop here or continue* for *For here or to go*. One point was the maximum score for any plausible response to option (c) or (d). The same was true for examples produced in a specific actual situational context. The definitions and examples were transcribed respectively, and two points were the maximum score for each item if learners received one point for a plausible definition and another for a plausible example. In the meanwhile, the mixed coding for further analysis was indicated in Table 4.3.

To be specific, Level 1 was composed of choosing both a & b options, together with wrong-answer and no-response options. For example, even if choosing option c, the response *You are all ready to leave* provided by one learner was also categorized

Table 4.3 Evaluation criteria for routine comprehension

Level	Rating criteria/score
1. No PC or ASC knowledge	choosing option (a) & (b) + total wrong answers or no response of (c) & (d), 0'
2. Plausible PC knowledge	choosing option (c) & (d) with explaining accurate definitions, 1'
3. Plausible ASC knowledge	choosing option (d) with raising a proper example, 1'
4. Plausible interplay of PC & ASC knowledge	choosing option (d) combined with a correct definition and a suitable example, 2'

Note ASC, actual situational context; PC, prior context

as Level 1 due to the erroneous statement. In addition, the correct definition that has been mentioned both in option c (*Here it is* for the target expression *All yours*) and d (*Do you want to eat food in the restaurant or take it away?* for the definition of *For here or to go*, together with the raised examples *Here is your coffee. For here or to go?*) belonged to Level 2. With respect to Level 3, one learner selected option d and only give the explanation to the definition of *Do you think you can make it?* as *Can you do it successfully?* but with no example. Such a situation should also be attributed to Level 3. While as to Level 4, see the above responses *Here is your coffee. For here or to go?*

4.4.3 Coding for Routine Recognition and Perception

Because the answers to the recognition task were relatively uniform, 1 point was awarded for one acceptable option based on their precise prior context knowledge in each scenario, for a total of 9 points. For example, the learner could only receive one point if he/she selected the target expression *Here you go* in Item 2; otherwise, the score would be zero.

Similarly, the criteria for routine perception were divided into two sections: learners' pragmatic awareness (up to 1 point for answering Yes, and 0 for No) and their adequate prior context knowledge. Each condition for their functional usage may be precisely targeted (1 point for one functional use, whereas one statement has two functional connotations, 0.5 points for each), with a total of 3 points for one pair and 15 points for the whole task. Here is an example,

No.

'Nice to meet you' is more formal;

'Nice to see you' is used when we say goodbye to somebody.

The answer *No* means actual differentiation, receiving 1 point. The functional usage of *say farewell to someone* received 0.5 points for not saying *whatever sort of*

person, known or unfamiliar s/he once met for the first/second time. Furthermore, the formal response to the use condition of *Nice to meet you* was completely incorrect, gaining 0 and finally 1.5 points in total.

4.4.4 Statistical Methods in Data Analysis

The independent t-tests with effect sizes (Cohen's d) were used to examine the impact of influencing variables on distinct components of routine task modalities in response to routine production, recognition, and perception. As Cohen (1988) elaborated, $0.2 < \text{Cohen's } d < 0.3$, insignificant effect size; Cohen's d approximately 0.5, medium effect size; Cohen's $d > 0.8$, large effect size. To evaluate group differences and the influence of variables in response to routine comprehension, McNemar chi-square and Mann-Whitney U tests were used sequentially. SPSS 23.0 was used for all data analysis procedures.

4.5 Verification for Inter-Rater Reliability

Inter-rater reliability investigates the extent to which different raters interpret the same set of data in the same way (Mackey & Gass, 2005), even with no specific rules established in the SLA presented. Nonetheless, inter-rater reliability is widely recognized as assessment indication for "checking the consistency and accuracy of coding" (Ren, 2014: 95), mainly "when high-inference categories are involved" (Kasper, 1998: 360). To establish inter-rater reliability, the researcher coded the rating criteria for each task and had them checked by another two experienced learners (a male doctor and a female master). In order to calculate inter-rater reliability, the researcher used the random sampling method to pick out 15% of responses in each task in total. The two raters then classified and scored these quantitatively filtered oral responses for appropriateness using the coding schemes developed in the prior chapter. Cohen's kappa was later used to determine if inter-rater dependability can reach an acceptable ideal level if its value exceeded 0.8. (Mackey & Gass, 2005).

Throughout every routine task modality, all Cohen's kappa values were more than 0.8. It is thus encouraging to see that the inter-rater reliability of all coding schemes for evaluation criteria was rated as excellent, with all reported values significantly exceeding 0.8 in the present study.

4.6 Ethical Considerations

All data collection occurred following approval from the East China Normal University Ethics Committee and all subjects who agreed to participate in this research. At

the same time, all participants were informed that this was a totally anonymous study. The data collected would be used solely for scientific research and statistical analysis and would not be exploited for any other purposes. You have the right to terminate the investigation at any moment throughout the test. When you started answering the questions, it shows you were fully informed and consented to participate in the study.

Concerning confidentiality, all data were transcribed anonymously. All participants were divided into three groups based on proficiency levels with or without international exposure, and no personal information was collected. The cases of relevant individuals were additionally tagged with group and serial number in the further analysis. To avoid data loss, the raw data and subsequent electronic transcripts must be held under absolute confidentiality.

4.7 Summary

This chapter described the entire design for the holistic study, including descriptive information on the three groups of participants (Sect. 4.1). This part also included an instrumentation summary (Sect. 4.2), as well as detailed data collecting protocols (4.3) and coding for each routine task (4.4). Sections 4.5 and 4.6 introduced the verification for inter-rater reliability and ethical problems in the present study, respectively. The outcomes of the exploration into each research question across each task modality will be analyzed in the next chapter.

References

- Bardovi-Harlig, K. (2014). Awareness of meaning of conventional expressions in second-language pragmatics. *Language Awareness*, 23(1–2), 41–56.
- Bardovi-Harlig, K. (2019). Routines in L2 pragmatics research. In N. Taguchi (Ed.), *Handbook of SLA and pragmatics* (pp. 47–62). Routledge.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Félix-Brasdefer, J. C. (2010). Data collection methods in speech act performance: DCTs, role plays, and verbal reports. In A. Martínez-Flor & E. Uso-Juan (Eds.), *Speech act performance: Theoretical, empirical and methodological issues* (pp. 41–56). John Benjamins.
- Halenko, N. (2018). Using computer-assisted language learning (CALL) tools to enhance output practice. In C. Jones (Ed.), *Practice in second language learning* (pp. 137–163). Cambridge University Press.
- Kasper, G. (1998). Analysing verbal protocols. *TESOL Quarterly*, 32(2), 358–362.
- Kasper, G. (2008). Data collection in pragmatics research. In H. Spencer-Oatey (Ed.), *Culturally speaking* (pp. 279–303). Continuum.
- Mackey, A., & Gass, S. (2005). *Second language research: Methodology and design*. Lawrence Erlbaum.
- Ren, W. (2014). A longitudinal investigation into L2 learners' cognitive processes during study abroad. *Applied Linguistics*, 35(5), 575–594.

- Ren, W. (2015). *L2 pragmatic development in study abroad contexts*. Peter Lang.
- Roever, C. (2005). *Testing ESL pragmatics: Development and validation of a web-based assessment battery*. Peter Lang.
- Roever, C. (2012). What learners get for free: Learning of routine formulae in ESL and EFL environments. *ELT Journal*, 66, 10–21.
- Taguchi, N., & Roever, C. (2017). *Second language pragmatics*. Oxford University Press.
- Wang, Y. (2022). The role of prior and actual situational context in conversational routines produced by Chinese learners of English. *Pragmatics & Society*, 13(4), 587–606.