

The Effectiveness of the Domestic AI Model “Doubao” in Correcting Japanese Business Email Errors: An Empirical Study at Shanghai Polytechnic University

Zheng Lu*

School of Languages and Cultural Communication, Shanghai Polytechnic University, Shanghai
201209, China

*Corresponding author

Abstract

This study looks at some persistent headaches in teaching Japanese business email writing—specifically, why students keep struggling with Keigo (honorific language) and the unspoken rules of business correspondence. Lately, AI tools like Doubao have started showing up in language classrooms, and they’re surprisingly good at giving students fast, tailored feedback. So I wanted to see: can AI actually make business Japanese writing instruction work better? I ran a quasi-experiment at Shanghai Polytechnic University. Two groups of undergrads, same course, same material. One group got to use AI to revise their email drafts; the other stuck with traditional teacher feedback only. I collected writing samples from both groups and compared error rates on keigo and sentence structure. The results? The AI-assisted group made noticeably fewer mistakes. But here’s the interesting part: it wasn’t the AI doing all the work. Students still needed their teacher’s input—the sweet spot seemed to be AI suggestions plus human guidance, layered together. They picked up on standard business conventions faster, wrote more confidently, and stopped making those awkward pragmatic errors that usually take months to unlearn.

Keywords

AI-assisted learning, Japanese business correspondence, writing instruction, Keigo, large language models.

1. Introduction

With China–Japan trade ties deepening, Japanese business writing has become a practical must-have for Chinese majors—yet classroom instruction hasn’t quite caught up. Business emails are much different from daily Japanese communication. They require strict rules for Keigo, fixed sentences and standard layouts (Chen, 2020; Qu, 2017). Past studies show that Chinese learners often rely on only one polite expression in business scenarios. They also make many honorific mistakes (Wu, 2015). Many students, for example, confuse respectful and humble language in business writing, or produce double honorifics (Jian, 2015). Chinese learners are influenced by their mother tongue, and Japanese honorifics are complicated, so students often make pragmatic and formatting errors.

Many studies have noted these learning difficulties. But university-level Japanese teaching in China still depends a lot on pattern memorization and teacher-led error correction. Students can hardly get fast feedback on pragmatic use. This teaching method works for grammar knowledge. Yet it gives students few chances to get quick, personal feedback for pragmatic skills. So many students finish their studies with theoretical knowledge only. They cannot deal with real-life business communication well. This gap calls for new approaches to business writing instruction. Large Language Models including ChatGPT and the home-grown Doubao have

brought new chances for language teaching. Unlike older CALL tools, today's large language models actually process language in ways that feel surprisingly human. They don't just flag errors—they explain why something is wrong, instantly (Kasneci et al., 2023). For Japanese learners, that's a big deal. Keigo, which most students dread, becomes less intimidating when you can practice it twenty times in a row without annoying a real person. The AI becomes a patient, always-available stand-in for a tutor who never gets tired of repeating the same rule.

AI writing tools have gotten plenty of attention in English composition classrooms—there's no shortage of studies there. But Japanese business correspondence? That's a different story. Empirical work on this specific context is almost nowhere to be found, which is odd when you think about it: business emails demand a level of sociolinguistic appropriateness that general English essays simply don't. Few studies have tested how well AI tools work in this specific learning scenario. Also, not enough research has focused on domestic Chinese AI models. Therefore, it is unclear how these models perform in business Japanese writing education.

This study tries to fill this gap. It looks at how well the Chinese AI model Doubao fixes errors in Chinese university students' Japanese business emails. This work tries to find answers to the research questions below:

RQ1: To what extent does Doubao reduce students' error rates in Keigo, sentence structure, and formatting, compared with traditional teacher feedback?

RQ2: What are students' perceptions of the usability and effectiveness of AI-assisted feedback in Japanese business email writing?

2. Methodology

2.1. Participants

This study had 40 participants. They were third-year students taking Japanese as their minor at Shanghai Polytechnic University. All students took the Japanese Foreign Trade Correspondence Writing elective course in the Spring Semester of 2025. Participants were divided into two groups for experimental control: the Experimental Group (EG, N=20) and the Control Group (CG, N=20). Students in both groups had similar Japanese proficiency, approximately at the JLPT N2 level. The course was taught by the researcher, ensuring instructional consistency across both groups.

2.2. Instruments and AI Tools

The primary instrument used in this research was Doubao, a Chinese domestic large language model developed by ByteDance (2025 version). All participants were Chinese undergraduates, so Doubao made sense as the tool of choice—it handles Chinese language nuances better than most competitors, and students could actually access it on campus without VPN hassles. To get useful feedback from the AI, I gave participants a specific prompt to paste in when submitting each draft:

You're an expert in Japanese business communication. A Chinese undergrad just wrote this email—your job is to find the Keigo mistakes, sentence structure problems, and formatting issues. Fix them, and explain in plain English why each change matters.

After the intervention, I also gave students a short questionnaire I'd put together myself. It used a standard 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree). The questions covered three things: whether the AI tool felt easy to use, whether its feedback actually helped them write more accurately, and whether they walked away satisfied with the whole experience.

2.3. Procedure

I ran the experiment across 8 weeks, fitting it into the regular course schedule so students didn't need extra sessions outside class.

Phase 1: Pre-test (Week 1). Both groups were assigned to write a " Inquiry about Product Prices " email. The drafts were collected and scored with a standard rubric for accuracy and appropriateness.

Phase 2: Intervention (Weeks 2–7). During these six weeks, both groups attended lectures on Japanese business email writing principles. However, their writing practice differed:

Control Group (CG): Students wrote weekly email drafts (e.g., Inquiry about Product Prices, Reply to a Complaint), and only got instructor feedback in the following class. They revised their work based on the teacher's handwritten comments.

Experimental Group (EG): Students wrote identical weekly drafts. Before receiving teacher feedback, they were required to input their drafts into Doubao using the designated prompt. They were instructed to read the AI-generated corrections and explanations in English, analyze the suggested changes, and revise their drafts accordingly. The AI-assisted revised draft was then submitted to the instructor for final evaluation.

Phase 3: Post-test and Questionnaire (Week 8). Both groups completed a post-test on a " Reply to a Complaint " email. Immediately after, the EG completed the perception questionnaire.

Data Collection and Analysis

The primary data source comprised the pre-test and post-test scripts. Two raters, including the instructor and a colleague blinded to the group allocations, assessed the emails. The error rate was calculated using the formula: $\text{Error Rate} = (\text{Number of Errors} / \text{Total Number of Clauses}) \times 100\%$. I sorted errors into three buckets: Keigo misuse (basically getting honorifics wrong), structural problems like grammar and syntax slip-ups, and formatting that didn't follow business norms. For the numbers—questionnaire responses and error rates—I ran everything through SPSS 26.0. The key comparison was an independent-samples t-test between the experimental group and the control group, and I set the significance threshold at $p < 0.05$.

3. Results

3.1. Error Rate Comparison

In Week 1, I gave both groups the same pre-test and found they were basically on par. The EG's starting error rate averaged 27.1% (SD = 4.2), while the CG sat at 26.8% (SD = 4.5). That difference is meaningless— $t(38) = 0.31$, $p = 0.76$, Cohen's $d = 0.08$. In plain terms: they started at the same place.

Table 1. Comparison of Error Rates Between Experimental Group (EG) and Control Group (CG) in the Post-test

Error Type	EG(n=20)	CG(n=20)	<i>t</i>	<i>p</i>
Keigo misuse	4.2%	10.5%	4.35	<0.001
Structural errors	5.1%	11.2%	3.98	<0.01
Formatting errors	3.2%	6.6%	2.87	<0.05
Total error rate	12.5%	28.3%	5.12	<0.001

Table 1 tells the story pretty clearly: the EG's total error rate dropped to 12.5% (SD = 3.1), while the CG barely budged at 28.3% (SD = 4.7), $t(38) = 5.12$, $p < 0.001$. The gap is real. And when you look at Keigo specifically, the drop is even more dramatic—EG students went from 10.5% down to 4.2%, basically cutting their honorific mistakes in half, $t(38) = 4.35$, $p < 0.001$.

3.1.1. Sample Error Analysis

To give a sense of what these errors actually looked like and how the AI fixed them, I pulled six examples from the EG students' drafts. For each one, I show the original student version, its Romanization, what the AI changed it to, and a short explanation of why.

Example 1: When students pick the wrong corporate title

	Content
Student draft (Japanese)	株式会社鈴木自動車殿
Romanization	<i>Kabushiki gaisha Suzuki Jidōsha dono</i>
AI correction	株式会社鈴木自動車御中
Romanization	<i>Kabushiki gaisha Suzuki Jidōsha Onchū</i>
Error type	Misuse of address suffix for a corporation (should be 御中, not 殿)
Explanation	dono is appropriate for individuals in formal internal correspondence, whereas Onchū is the standard suffix when addressing a company as a whole in business emails.

Example 2: Double Honorific in Personal Address

	Content
Student draft (Japanese)	佐藤先生様
Romanization	<i>Satō sensei-sama</i>
AI correction	佐藤先生 (or 佐藤様)
Romanization	<i>Satō sensei (or Satō sama)</i>
Error type	Double honorific
Explanation	Adding sama after sensei doubles up on honorifics—something Japanese business writing doesn't allow. The AI caught it and suggested either Satō sensei or Satō sama, stripping it back to one suffix and keeping it within normal business register.

Example 3: Incorrect Form of Request in a Price Inquiry Email

	Content
Student draft (Japanese)	お返事いただきます。
Romanization	<i>O-henji itadakimasu.</i>
AI correction	お返事いただけますか。
Romanization	<i>O-henji itadakemasu ka.</i>
Error type	Incorrect form of request(statement vs. question)
Explanation	In a price inquiry email, the student used "O-henji itadakimasu" (I will receive your reply) as a closing line, which is grammatically a statement rather than a polite request. The AI corrected it to "O-henji itadakemasu ka" (Could you please reply?), which appropriately uses the potential form + question particle to express a polite request.

Example 4: Misuse of Respectful Form for Speaker's Own Action

	Content
Student draft (Japanese)	このたび商用で貴地に出張される際、貴社にお伺いしたく存じます。
Romanization	<i>Kono tabi shōyō de kichi ni shucchō sareru sai, kisha ni oukagai shitaku zonzimasu.</i>
AI correction	このたび商用で貴地に出張する際、貴社にお伺いしたく存じます。
Romanization	<i>Kono tabi shōyō de kichi ni shucchō suru sai, kisha ni oukagai shitaku zonzimasu.</i>
Error type	Misplaced honorific (respectful form used for speaker's own action)
Explanation	Here's where it gets interesting. A student wrote "shucchō" (a speaker's own action) with the respectful auxiliary -rareru (sareru), which is just wrong—you don't elevate your own actions in Japanese business writing. The AI caught the mistake and stripped the honorific (sareru → suru), which fixed the grammar but missed the bigger picture. From a business etiquette standpoint, the better move would've been itasu—the humble form—because in Japanese business correspondence, everything you do should be expressed with humility, not just stripped of honorifics. So the AI got halfway there: it spotted the pragmatic mismatch and corrected the grammar, but it didn't push to the next level of appropriateness. That gap is exactly where a teacher needs to step in.

Example 5: Inappropriate Person Reference in Business Email

	Content
Student draft (Japanese)	あなたの注文は準備中です。
Romanization	<i>Anata no chūmon wa junbi-chū desu.</i>
AI correction	ご注文は準備中です。
Romanization	<i>Go-chūmon wa junbi-chū desu.</i>
Error type	Inappropriate person reference (using anata for a client)
Explanation	One student opened with anata (you) to address the client directly—a move that falls flat in Japanese business writing because it sounds way too casual for the context. The AI swapped anata no kaisha ("your company") for kisha (貴社, "your honorable company"), which is the standard humble way to refer to a client's organization. This kind of slip shows up a lot with Chinese learners: they translate "贵公司" word-for-word into "your company" and stop there, missing the whole honorific layer Japanese demands. The fix here proves the AI can catch pragmatic problems in how you refer to people—not just grammar, but actual social register.

Example 6: Inappropriate Directness in Email Body

	Content
Student draft (Japanese)	メールを送りました。確認してください。
Romanization	<i>Mēru o okurimashita. Kakunin shite kudasai.</i>
AI correction	メールを送信いたしました。ご確認いただけますでしょうか。
Romanization	<i>Mēru o sōshin itashimashita. Go-kakunin itadakemasu deshō ka.</i>
Error type	Inappropriate directness (imperative instead of polite request)
Explanation	A student asked for confirmation using a plain polite form (okurimashita) plus a direct kudasai—fine for casual emails, but way too blunt for business correspondence. The AI fixed both: first, it upgraded okuru to the humble form sōshin itashimashita (送信いたしました); second, it swapped the bare kudasai for the much softer go-kakunin itadakemasu deshō ka (ご確認いただけますでしょうか). This is a good example of the AI doing more than just catching grammar mistakes—it actually nudged the whole email up to a proper business register.

4. Discussion

So what did we actually find? Using Doubao for feedback made a real difference. Students in the EG ended up with a total error rate of 12.5%, compared to 28.3% in the CG—basically cut their mistakes in half. This difference is statistically significant ($t(38) = 5.12, p < 0.001$). This lines up with what Wu (2015) and Jian (2015) found—Chinese learners keep tripping over the same things: Keigo and formatting. And it backs up Kasneci et al. (2023) on a bigger point: AI tools, when used right, actually help students stop making these mistakes.

Looking at the six error examples more closely, the picture gets more interesting. Some cases show Doubao doing exactly what it's supposed to do. In Example 3 (wrong request form) and Example 6 (too-direct tone), the AI caught pragmatically awkward phrasing and rewrote it into proper business language—no human help needed. Same goes for Example 5 (messing up how you refer to people) and Example 1 (stacking honorifics on a company name): it spotted mistakes that Chinese learners typically don't even notice, like context-specific honorific errors and formatting slip-ups. But there's a ceiling. Example 4 is the clearest case—a student put a respectful auxiliary (sareru) on their own action (shucchō), and Doubao fixed the grammar by changing it to suru. Fine, but that's where it stopped. It didn't take the extra step of upgrading to the humble form itasu, which is what business etiquette actually demands. The pattern is clear: the AI prioritizes grammatical correctness and stops there, missing opportunities to push students toward higher pragmatic appropriateness. That's where teachers come in—they need to catch what the AI leaves behind and walk students through those finer points. Example 2 (double honorific) tells a similar story. The AI found the redundant suffix and stripped it out, but when it came to suggesting the best alternative, it fell short without a teacher filling the gap. What this all points to is pretty straightforward: Doubao works best as a first-pass tool, not a replacement for actual teaching. It's genuinely good at catching surface-level problems—formatting slip-ups, basic Keigo mix-ups, overly direct phrasing—and it does this fast and consistently. In classes where you've got thirty-plus students and maybe two contact hours a week, that kind of coverage is something traditional teacher correction simply can't match. No human can give timely, detailed feedback on every draft. But—and this is the important part—the AI starts to falter when language use gets socially complex. It doesn't reliably pick the right register for a given situation: plain form versus humble versus respectful, that whole layer of

pragmatic choice that depends on who's writing to whom and why. These are exactly the places where a teacher needs to step in. So the practical takeaway is a two-stage model. Students run their drafts through AI first and fix what they can on their own—most of the mechanical errors disappear at this stage. Then the teacher comes in, not to redo the basics but to catch what the AI left behind: the pragmatic misfires, the register mismatches, the subtle stuff that actually separates a competent business email from a good one. It's not AI versus teachers. It's AI first, teachers second, and students who actually learn something in between.

5. Limitations and Future Research

A few caveats, naturally. The sample was small—40 students from one university—so I'm not pretending these results generalize everywhere. Eight weeks is also a short window; whether the gains stick around six months later is anyone's guess without a longer follow-up. Then there's the tool itself: I only used Doubao. Throwing other LLMs like ChatGPT, Claude, or ERNIE Bot into the mix would make for a much more useful comparison—which one actually fits this kind of language task best is still an open question. And on the analysis side, I didn't comb through every single student submission; I picked representative error examples instead. A full corpus-based approach would've given richer detail, but that's a different project.

6. Conclusion

Here's the bottom line: Doubao, a domestic AI model, actually works for Japanese business email correction. Students using AI feedback made noticeably fewer mistakes across the board, and the biggest win was in Keigo—exactly where Chinese learners need the most help. Every error category showed improvement in the EG compared to the CG, with honorific accuracy showing the widest gap. That said, the tool isn't a complete package. It handles the mechanics well, but when writing decisions get socially nuanced, you still need a teacher in the room.

This study chips in something new to the AI-assisted language education conversation. Locally built LLMs like Doubao aren't just usable—they actually fit the specific needs of business Japanese learners in ways that older writing tools never could. The key advantage is practical: students can type in Japanese and get explanations back in English, which traditional tools simply don't do.

This study also makes a case for how AI and teachers can actually split the work. The AI's edge is obvious: it's instant, it doesn't get tired, and it gives the same quality feedback whether you have 15 students or 50. No instructor can match that kind of consistency, especially in a packed classroom. But Example 4 showed the ceiling—when it comes to subtle pragmatic choices, the AI just doesn't get there. That's not a flaw you can patch; it's baked into how these models work. So the answer isn't picking sides. Let the AI take the first pass on the mechanical stuff, and let teachers handle the layer that actually requires judgment: when to use humble, when to pull back, when "correct" isn't good enough. That division of labor is where the real value is.

Of course, this study isn't without its rough edges—40 students from one school, eight weeks, one tool. That's a narrow slice. But the results still hold up as real empirical evidence that AI has a place in business Japanese writing instruction. Where to go from here is obvious: get more institutions on board, run it longer, and test Doubao against other LLMs to see if the pattern holds. And as these models keep getting better, the smart move for teachers isn't to wait for perfection—it's to start using what works now, and keep learners' actual needs at the center of it.

References

- [1] Chen, Z. Y. (2020). Japanese honorifics and feminine language in business activities. *Journal of Japanese Language Study and Research*, 2, 9–16.
- [2] Qu, S. (2017). Use of Japanese honorifics in business correspondence and pedagogical reflections. *Industrial & Science Tribune*, 16, 181–182.
- [3] Wu, Y. X. (2015). Theoretical basis for problems in Japanese honorific acquisition among Chinese students. *Foreign Language Education*, 36, 57–60.
- [4] Jian, Z. Z. (2015). Key points and techniques in writing Japanese business documents. *Xueyuan*, 34, 74–75.
- [5] Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gül, C., Günther, U., Hüllermeier, E., Krusche, S., Kutrubes, M., Lüders, H., & Reichert, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- [6] Zhou, X. P. (2018). Features of honorific usage in business Japanese correspondence. *English Square*, 3, 67–69.

Declaration of Generative AI

I used the Chinese LLM Doubao (ByteDance, 2025 version) strictly as a tool in this study: it gave participants corrective feedback on their Japanese business email drafts, nothing more. It was not used to generate manuscript text, nor is it an author or reference source. All error counts, example selections, and analyses were produced by the researcher; model output was independently verified.