

TAMMY: Supporting EFL Translation Practice with an LLM-Powered Chatbot

Steve WOOLLASTON*, Brendan FLANAGAN, Patrick OCHEJA, Yiling DAI & Hiroaki OGATA

Graduate School of Informatics, Kyoto University, Japan

*s.m.woollaston@gmail.com

Abstract: Learning EFL through translation tasks is an effective language learning technique, but requires consistent practice and scaffolding. This study evaluates TAMMY, a prototype EFL chatbot designed for Japanese learners to practise English translation tasks. Using the extended Technology Acceptance Model, the study examines Tammy's usability, usefulness, and enjoyment. Response appropriateness and task success are also explored. Findings from a pilot study with Japanese university students indicate high usability and positive attitudes towards the chatbot. Tammy effectively provided accurate feedback in most tasks successfully guiding learners to an accurate translation, though improvements are needed in feedback clarity and conversational adaptability. Despite limitations, Tammy shows promise as a support tool for language learning, offering an engaging and non-judgmental platform for practising translation and enhancing English proficiency.

Keywords: Chatbot, conversational agents, EFL, translation, usability, chatlogs, exTAM, language learning, GenAI, LLM, TELL

1. Introduction

English as a Foreign Language (EFL) teaching and learning is utilising technology more and more (Zhang & Zou, 2022), particularly through Computer Assisted Language Learning (CALL) systems. These systems leverage the capabilities of hardware and software to enhance and augment language learning experiences and outcomes, offering tools that range from vocabulary drills such as Anki Flashcards (n.d.) to comprehensive language courses like Duolingo. Chatbots have emerged as a promising tool for language practice and instruction, providing an interactive and engaging dialogue-based platform for learners to gain exposure to natural conversations and improve their skills in the target language.

Chatbots, or conversational agents, have shown potential in various educational contexts due to their ability to simulate human-like interactions. These interactions can provide immediate feedback, adapt to the learner's needs and language proficiency, and offer a non-judgemental space for practising language skills and making mistakes. In learning and teaching EFL, chatbots are valuable tools for practising conversational English, vocabulary and grammar, and as this paper will propose - translation tasks. As recent research has demonstrated (Dagiliené, 2012), translation practice plays a crucial role in language learning as it reinforces vocabulary, syntax, and comprehension while making important connections with a learner's first language.

In this study, we introduce and evaluate an EFL translation practice chatbot named TAMMY (Translation Assistant for MMasterY). Tammy is designed to provide Japanese learners of English with practice translation tasks, and assist with the translation process by providing feedback and guidance. The system is evaluated using an extended Technology Acceptance Model (exTAM) and chatlog data. This exploratory study focuses on the following research questions:

- RQ1:** *How do learners perceive Ease of Use, Usefulness, and Enjoyment of the chatbot? What are their Attitudes and Intentions toward the chatbot?*
- RQ2:** *How valid are the chatbot's responses to the learner, and how effective is it in guiding learners to good translations?*

RQ3: *What are learners' perceptions of the strengths, weaknesses, and areas for improvement of the current chatbot?*

2. Related Work

2.1 Translation tasks for learning EFL

While translation tasks have waned in and out of popularity, research has consistently demonstrated their effectiveness in second language learning. From the Grammar-Translation Method and behaviourism's language drills to today's cognitive and communicative approaches, translation tasks have remained central to language teaching. (Antón Remírez & Lázaro Ibarrola, 2011). Translation tasks are effective language learning activities, particularly in terms of:

1. Vocabulary acquisition (Folse, 2004)
2. Grammar and syntax: Translation tasks provide numerous formulaic phrases and help learners understand and apply grammatical structures more accurately (Murtisari et al., 2020)
3. Reading comprehension: Translation exercises enhance reading comprehension by promoting close attention to details (Chang & Hsu, 2011)
4. Writing proficiency: Translation tasks improve writing by enhancing accuracy, cohesion, and coherence (Tavakoli et al., 2014) and reducing cognitive load by allowing idea development in L1 before translation. (Huh, 2001; Saleh et al., 2021).

2.2 Translation using CALL

Learners and teachers increasingly use technology, known as Computer Assisted Language Learning (CALL), in both formal and informal contexts (Levy & Stockwell, 2013). Numerous free and paid apps and tools are available for learning, practising language skills, and connecting with language partners (Bowker et al., 2008). Learning English is particularly popular due to its role as the lingua franca in international business and academia (Smith, 2005).

Several English Language Learning Apps (ELLA) and digital tools incorporate translation as a key component of their pedagogy, such as: a translation-based approach to teach new vocabulary and grammar in **Duolingo** (Loewen et al., 2019), translating words and phrases using mnemonics and repetition to reinforce knowledge in **Memrise** (Karjo & Andreani, 2018), conversation-based simulated dialogues for sentence translation in practical contexts in **Babbel** (Kessler et al., 2023), and the use of machine translation tools to learn languages by translating words, phrases, and sentences to and from their native language in **Google Translate** (Inderawati et al., 2023).

Several studies have examined learner perceptions of ELLA. Nami (2020) found that university students were sceptical about apps improving listening, speaking, and writing skills but had highly positive views on using them for vocabulary and mostly positive views on reading comprehension. Metruk (2021) found EFL learners favoured ELLA for enabling English practice anytime, anywhere, being easy, flexible, convenient, effective, and enjoyable. Learners noted ELLA was most helpful for developing vocabulary and least for speaking skills and confidence.

2.3 Chatbots for Language Learning

Chatbots gained prominence in user support for their ease of use, providing instant, 24/7 assistance. Research into response validity has improved their effectiveness and reliability (Casas et al., 2021). In classrooms, chatbots offer a non-judgmental learning environment, reduce teacher workload, and allow self-paced learning (Ait Baha et al., 2023; Cunningham-Nelson et al., 2019). However, issues like novelty effects and technological limitations remain (Huang et al., 2022)

Research consistently shows that chatbots significantly improve second language proficiency, especially with corrective feedback (Bibauw et al., 2022; Deng & Yu, 2023; Woollaston et al., 2024; Zhai, 2023). Achieving natural language skills requires extensive dialogue, which chatbots provide as an accessible alternative to proficient speakers, available anytime without fatigue. With the advent of Large Language Models (LLMs), chatbots are now more capable and less prone to errors, leading to a surge in their use in EFL contexts (Woollaston et al., 2024). Learners prefer chatbots for their non-judgmental nature, patience with repetitive errors, and broad conversational abilities (Tekwa, 2024). Chatbots offer a safe space to practise new phrases and structures without fear of judgement, reducing anxiety. However, challenges and ethical issues exist, such as the lack of real human interaction, hallucinations, factual errors, privacy concerns, and bias in training data. These issues are discussed comprehensively by Rebolledo Font de la Vall and Araya (2023).

ESL teachers generally find chatbots useful for providing feedback and facilitating student interaction in the target language (Chuah & Kabilan, 2021; Yang, 2022). In a study of 142 teachers, Chuah and Kabilan (2021) found that chatbots helped to create an engaging and active environment by “simulating human-like social interactions” (p. 233), leading to a greater level of social presence. Teachers appreciated the ease of use and integration with current systems (e.g. messaging apps) chatbots provide, and noted that additional training would be useful to effectively use chatbots in their teaching programmes.

3. Methodology

3.1 Proposed System

Building on the efficacy of translation tasks for language learning and advanced chatbot technologies, we propose Tammy, an EFL translation practice chatbot. Tammy will allow learners to practise translating sentences from Japanese to English, receive feedback, and engage in dialogue to clarify understanding and reinforce learning. Unlike existing educational chatbots, Tammy will focus exclusively on translation tasks and provide personalised scaffolding based on learners' responses and questions. This study aims to evaluate Tammy's perceived usability, enjoyment, and usefulness, assess the chatbot's responses to translation attempts and questions, and also gather learners' feedback on strengths, weaknesses, and areas for improvement.

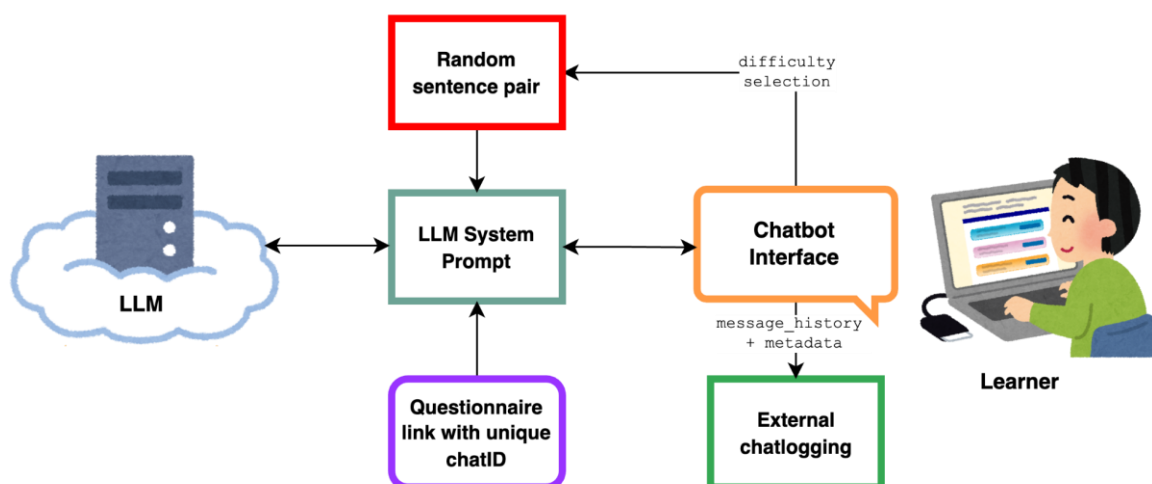


Figure 1. Tammy chatbot system overview.

3.2 Tammy System Overview

Tammy is a simple chatbot designed to facilitate translation practice from Japanese to English. Tammy is powered by gpt-3.5-turbo-0125 and was developed using Gradio for the UI (Figure 1). To avoid hallucination and encourage the LLM to provide more accurate and factual responses and mitigate hallucination, the temperature was set to 0.2. No other parameters were utilised (Watkins, 2023).

Tammy provides the following features:

- **Difficulty Selection:** Students can select sentences that match their proficiency
- **New Sentence Selection:** Request new random sentences at selected difficulty
- **Retry:** Where the response is unsatisfactory or unhelpful, learners can ask the chatbot to attempt a response again
- **Undo:** Allows users to delete the chatbot's response and their message to retry that part of the conversation
- **Examples:** Common questions and commands are available to facilitate ease of use
- **Language:** The chatbot defaults to English, but is also able to converse in Japanese.

3.3 Learning Content

For each difficulty level (*easy*, *intermediate*, *advanced*), twenty Japanese and English sentence pairs were created using ChatGPT-4, based on examples from the Dictionary of Japanese Grammar book series (Makino & Tsutsui, 1986, 1995, 2008). Native speakers of both languages reviewed the generated sentences for accuracy. A sample of each sentence pair is shown in Table 1. A random sentence pair was chosen for each new session (default Easy). Learners can randomly select a new sentence pair or change the sentence difficulty.

Table 1. *Sample of sentence pairs*

Difficulty	Japanese Sentence	English Sentence (model answer)
Easy	これはペンです。	This is a pen.
	猫が好きです。	I like cats.
Intermediate	私たちは映画を見た後でレストランに行きます。	We will go to a restaurant after watching the movie.
	彼女は猫と遊ぶのが好きです。	She likes playing with cats.
Advanced	その歴史的イベントは今日の私たちの社会に多大な影響を与えています。	That historical event has a great impact on our society today.
	この文化の違いは時にはコミュニケーションを難しくします。	Cultural differences sometimes make communication difficult.

The system prompt was refined through a process of iterative trial and error, enabling the chatbot to respond accurately to a broad range of learner inputs.

3.4 Data Collection

Participants were 12 university students (proficient Japanese speakers) from a top national university in Japan. In May 2024, they were invited to participate anonymously via an internal message board in the Learning and Educational Technologies Research Unit. A questionnaire was created using Google Forms and linked in the chat after learners either correctly translated a sentence or chose to stop. Each chat session was linked to a completed

questionnaire using a chat identification number. Statistics were calculated using JASP (JASP Team, 2024). Chatlogs were reviewed, and system-generated entries and six drive-by instances were removed. The feedback form URL was excluded for character counting. Descriptive statistics for the chatlogs were calculated using Google Sheets. The chatbot's responses were categorised as valid, partially valid, or invalid: valid responses were fully accurate and contextually appropriate, partially valid responses contained mostly correct information but also errors or irrelevant content, and invalid responses were incorrect or inappropriate. Examples are provided in the results section. The system prompt, questionnaire, and complete list of all 60 sentence pairs is available at: bit.ly/2024tammyICCE.

3.5 Measures

The exTAM instrument (extended Technology Acceptance Model), as developed by (Wu & Gao, 2011), was adapted for use evaluating a chatbot. It is an established model for measuring affective factors that influence learners' attitudes and behavioural intentions toward using interactive learning technologies (Granić & Marangunić, 2019). The questionnaire consists of 17 mixed questions where respondents indicate their agreement to statements from Strong Disagree to Strongly Agree on a seven-point likert scale. Scale items are averaged for each of the five factors: Perceived Ease of Use (PEOU), Perceived Usefulness (PU), Perceived Enjoyment (PE), Attitude (ATT), and Intention.

For qualitative analysis, three required open questions were also given. These questions align with the method utilised by Kim et al. (Kim et al., 2021):

1. *What did you like about Tammy?*
2. *What did you not like about Tammy?*
3. *What could be improved in Tammy?*

Responses were categorised, and patterns and themes described. Data on self-reported literacy proficiency in both Japanese and English was also collected (Beginner, Intermediate, Advanced, Native-level).

4. Results

4.1 Affective Factors - exTAM

Table 2 shows generally positive feedback across all factors of the exTAM. PEOU averaged 5.75, indicating high usability, with "Learning to use Tammy" scoring the highest at 6.00. PU was lower at 5.00, showing varied opinions, especially regarding productivity benefits. Users found the chatbot enjoyable, averaging 5.42 in PE, and displayed a favourable attitude (ATT) with an average of 6.11. Intention to use Tammy showed positive yet varied responses, with an average of 5.42, suggesting a readiness to recommend and continue using Tammy.

Table 2. *exTAM Descriptive Results (n=12)*

Factor	Reliability	Mean	SD
Perceived Ease of Use (PEOU):	0.76	5.75	0.87
Perceived Usefulness (PU):	0.88	5.00	1.11
Perceived Enjoyment (PE):	0.83	5.42	1.07
Attitude (ATT):	0.34	6.11	0.59
Intention:	0.46	5.42	0.81

The reliability of the Attitude (ATT) and Intention factors, as indicated by Cronbach's alpha, are notably low. This suggests that the internal consistency of the items measuring these factors is weak, and that the items may not be capturing a single construct. The small number of participants in this pilot study may also have a detrimental impact on the reliability. Despite the low reliability, the mean scores are relatively high, suggesting that participants generally

have a positive view towards using Tammy for learning English and plan to use the system in the future.

4.2 Chatlogs

All messages between the chatbot and Tammy's response were recorded with corresponding metadata (timestamp, role, difficulty, chat ID, English and Japanese sentence). There were a total of 482 messages exchanged between the chatbot and learners in 17 unique user sessions and 40 translation tasks (Japanese-English sentence pairs). Of the messages, 204, 104, and 174 messages were at Easy, Intermediate, and Advanced difficulty respectively.

Table 3. *Length of each session*

	Mean	SD	Median	Min	Max
Number of messages	23.52	17.31	20	8	76
Duration (min:sec)	10:37	12:50	04:20	01:02	47:08

As shown in Table 3, the shortest session was 1:02s and successfully completed two translation tasks: one easy and one intermediate sentence. The longest session (47 mins) successfully completed four translation tasks (two easy, one intermediate and advanced), had one chatbot error and abandoned one sentence. There were also large gaps in their responses, indicating the learner may have been distracted while using the chatbot. Figure 2 shows the message lengths, measured in characters between the chatbot and the learners. Word counts were not calculated as it does not make sense to count words in the Japanese language. The median and mean message length sent by the learner is notably lower when compared to the chatbot. The whiskers, representing the range excluding outliers (data points that fall outside 1.5 times the interquartile range), show that most of the learner's messages are clustered close to the median, with a slight skew towards shorter messages.

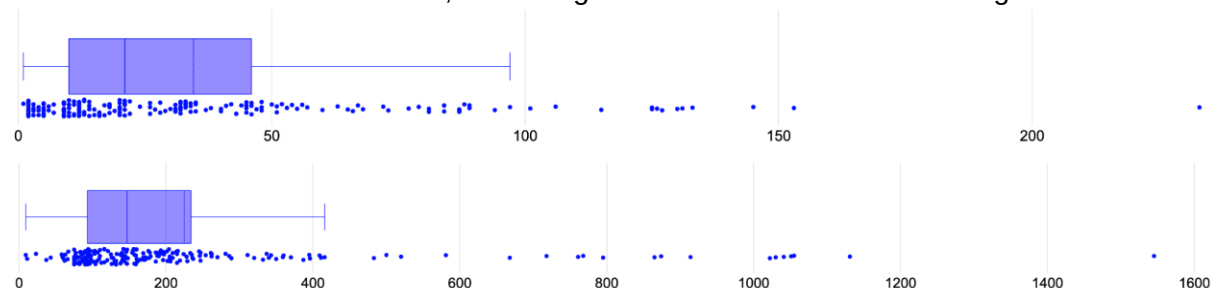


Figure 2. Box plots of learner (top) and chatbot (bottom) message lengths (characters).

The median message length for the chatbot is significantly higher than that of the EFL learner. Further, the interquartile range is wider for the chatbot, implying a greater variability in message lengths and perhaps highlighting the adaptive nature of chatbot interactions in supporting language learners. In the future, a limit on the length of chatbot responses will be implemented.

4.2.1 Chatbot Response Evaluation

Out of 197 messages, the chatbot responded appropriately 78% ($n=153$) of the time, indicating that it generally delivers accurate and relevant responses. 9% ($n=18$) of the responses were considered partially valid, where the message was not entirely accurate or fully appropriate. Invalid responses made up 13% ($n=26$) of the total responses, due to their unhelpful or irrelevant nature. Interestingly, of the invalid and partially valid responses, many occurred after the sentence had been correctly translated and the user subsequently continued the conversation. The chatbot was not able to adapt to this new context, as illustrated in Table 4.

Table 4. Sample chatlog validity evaluation

Learner	Tammy Response	Validity	Justification
Hi!	Hello! Can you please translate the Japanese sentence "猫が好きです。" to English?	Valid	Proper greeting and clear instruction.
I like cats.	"Great job! Your translation is correct!"	Valid	Accurate confirmation and encouraging feedback.
I like dogs, too.	The translation is almost correct. Remember to include the word "too" in your sentence to match the meaning of the original Japanese sentence. Can you try again?	Invalid	Incorrect feedback; no original Japanese sentence mentioned about dogs.
How about you, Tammy? Do you like cats or dogs?	I appreciate your question! Let's focus on the translation first. Can you please try translating the Japanese sentence "猫が好きです。" to English?	Partially valid	Redirects to translation task, but does not engage with the personal question appropriately. Translation already completed.
I like cats.	"Great job! Your translation is correct!"	Valid	Accurate confirmation and feedback request repeated.

Of the 40 total sessions, only three were derailed by invalid responses by the chatbot. Four were abandoned for unknown reasons. 82.5% ($n=33$ translation attempts) were either adequate ($n=4$) or ideal ($n=28$) in guiding the learner to the correct translation.

4.2.2 Open Question Analysis

The feedback for Tammy revealed several key themes. Participants appreciated Tammy's friendliness and frank demeanour, which made interaction enjoyable and engaging. *"Tammy is frank, so I could enjoy using it."* Effective feedback was highly valued, as learners found it convincing and helpful in identifying and rectifying their errors. *"It was good for me that my mistakes were subdivided and what I don't know became clear gradually."* The conversational approach to learning was particularly praised, with learners noting that it made the learning more interactive and immediate. *"The idea of learning English with conversations was very attractive for me."* Persistence was another positive aspect, with Tammy consistently helping users reach a correct translation. *"Tammy attempted many times forever until I reached the goal."*

However, there were some areas of dissatisfaction where improvements are required. Some participants found the feedback vague or unclear, especially when it came to subtle nuances. *"Sometimes I did not get why my sentence was needed correction, e.g., 'The boy' and 'That boy' in a short sentence could refer to the same meaning."* Usability issues with the chatbot interface were also noted. At the advanced difficulty level, some users found the feedback repetitive or frustrating. *"What Tammy suggested was completely the same as the answer."* Additionally, some cultural aspects were overlooked for Japanese learners. *"...starting with 'Hello' is not very natural for Japanese people."* (Translated). Participants suggest expanding Tammy's focus to include other languages and allowing for more customisation in the specificity of hints and feedback.

5. Discussion and conclusion

This study proposed an EFL translation practice chatbot for learners to exercise and improve

their Japanese to English translation skills. To assess the usability and response validity of a chatbot, a simple prototype was created and piloted with a small number of Japanese university students.

RQ1: How do learners perceive Ease of Use, Usefulness, and Enjoyment of the chatbot? What are their Attitudes and Intentions toward the chatbot?

Based on the exTAM instrument given to participants after their interaction with Tammy, learners generally perceive the chatbot positively across all dimensions. The PEOU factor especially received the highest average score, suggesting that learners found the chatbot highly usable. This is consistent with previous research which has demonstrated that chatbots are an effective tool for facilitating learner-computer interactions in educational settings (Ren et al., 2019). While still positive, the lowest factor, PU, garnered more varied responses which suggests learners have mixed views on the utility and effectiveness of Tammy's educational benefits. This is despite the notably high positive Attitude (ATT) toward Tammy and Intention to utilise in the future, which suggests a strong enthusiastic sentiment toward the chatbot. Responses to the open questions support these findings, such as the feedback that praises Tammy's friendliness and the enjoyment felt when using the chatbot.

RQ2: How valid are the chatbot's responses to the learner, and how effective is it in guiding learners to good translations?

Tammy is generally effective at guiding learners through translation tasks, providing acceptable responses 87% of the time. Leveraging large language models (LLMs) for interpreting learners' questions and attempts proves effective in most cases, offering a significant improvement over traditional rule-based chatbots. Issues did arise when learners engaged in dialogue with the chatbot beyond the translation task, suggesting Tammy is limited in its adaptability. Tammy successfully helped learners complete over 82% of the translation tasks, showing that with appropriate scaffolding and persistence, learners can overcome the challenges inherent in translation. As one participant noted, "Tammy attempted many times forever until I reached the goal."

RQ3: What are learners' perceptions of the strengths, weaknesses, and areas for improvement of the current chatbot?

Learners highlighted several strengths and areas for improvement in their open text responses. Learners acknowledged the mostly effective feedback given by the chatbot, noting its friendliness and willingness to converse (even beyond the translation task). Learners also appreciated the immediacy of the feedback provided, which has been shown to enhance engagement and retention (Epstein et al., 2002).

Areas for improvement suggested by learners include enhancing Tammy's feedback clarity and consistency. Due to the stochastic nature of LLMs, at times the chatbot would respond with unhelpful advice or ask for another translation attempt where the provided translation was already accurate; these errors may be mitigated with better prompting. Learners also pointed out that Tammy's struggled to deal with conversational shifts beyond the translation task during extended interactions, and that the chatbot's contextual awareness needed development. Several usability, cultural appropriateness, and guidance issues were also raised which should be addressed in future versions of the chatbot.

The limitations of this study include a small sample size, with 12 participants completing 17 user sessions and 40 translation tasks in total. This may not fully represent the diversity of learner interactions. The novelty effect may have influenced learners' perceptions, as noted in previous chatbot research (Huang et al., 2022). The Tammy system itself also has several limitations. The current chatbot only supports translation from Japanese to English; expanding this to bidirectional translation would be beneficial. Future research should involve a larger, more diverse sample and incorporate objective measures of learning outcomes to extend the findings on usability and enjoyment and better understand the impact on learning outcomes.

This pilot study highlights the potential of the LLM-powered chatbot, Tammy, to enhance EFL

translation practice through personalised feedback and engaging interaction. Positive user perceptions and generally accurate feedback suggest Tammy has potential as a supplementary language learning tool. Furthermore, AI-powered chatbots like Tammy could foster learner autonomy, personalise support, and increase accessibility to language learning resources. Future research should investigate their long-term learning impact and explore their potential for addressing diverse learning needs and contexts.

Acknowledgements

Thank you to the participants who took the time to try Tammy and share their valuable feedback. Special thanks to Taisei Yamauchi for assisting with translations. This work was partly supported by JSPS Grant-in-Aid for Scientific Research (B) JP20H01722 and JP23H01001, (Exploratory) JP21K19824, (A) JP23H00505, and NEDO JPNP20006.

References

- Ait Baha, T., El Hajji, M., Es-Saady, Y., & Fadili, H. (2023). The impact of educational chatbot on student learning experience. *Education and Information Technologies*.
- Anki. (n.d.). *Anki - Powerful, Intelligent Flashcards*. Retrieved 22 May 2024, from <https://apps.ankiweb.net/>
- Antón Remírez, S., & Lázaro Ibarrola, A. (2011). Lost in translation? Translation as a valid tool for the EFL classroom. *Huarte de San Juan*.
- Ashouri, A., & Fotovatnia, Z. (2010). The effect of individual differences on learners' translation belief in EFL learning. *English Language Teaching*, 3, 228–236.
- Bibauw, S., Van den Noortgate, W., François, T., & Desmet, P. (2022). Dialogue systems for language learning: A meta-analysis. *Language Learning & Technology*, 26(1).
- Bowker, L., McBride, C., & Marshman, E. (2008). Getting more than you paid for? Considerations in integrating free and low-cost technologies into translator training programs. *Redit: Revista Electrónica de Didáctica de La Traducción Y La Interpretación*, ISSN 1989-4376, N^o. 1, 2008, Pags. 26-47, 26–47.
- Casas, J., Tricot, M.-O., Abou Khaled, O., Mugellini, E., & Cudré-Mauroux, P. (2021). Trends & Methods in Chatbot Evaluation. *Companion Publication of the 2020 International Conference on Multimodal Interaction*, 280–286.
- Chang, C.-K., & Hsu, C.-K. (2011). A mobile-assisted synchronously collaborative translation–annotation system for English as a foreign language (EFL) reading comprehension. *Computer Assisted Language Learning*, 24(2), 155–180.
- Chuah, K., & Kabilan, M. (2021). Teachers' views on the use of chatbots to support English language teaching in a mobile environment. *International Journal of Emerging Technologies in Learning (iJET)*, 16.
- Cunningham-Nelson, S., Boles, W., Trouton, L., & Margerison, E. (2019). *A review of chatbots in education: Practical steps forward* (pp. 299–306). Engineers Australia.
- Dagilienė, I. (2012). Translation as a Learning Method in English Language Teaching. *Studies about Languages*, 21, 124–129.
- Deng, X., & Yu, Z. (2023). A Meta-Analysis and Systematic Review of the Effect of Chatbot Technology Use in Sustainable Education. *Sustainability: Science Practice and Policy*, 15(4), 2940.
- Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., & Brosvic, G. M. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *The Psychological Record*, 52, 187–201.
- Folse, K. S. (2004). Myths about teaching and learning second language vocabulary: What recent research says. *TESL Reporter*, 37, 1–13.
- Granić, A., & Marangunić, N. (2019). Technology acceptance model in educational context: A systematic literature review. *British Journal of Educational Technology: Journal of the Council for Educational Technology*, 50(5), 2572–2593.
- Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer*

- Assisted Learning*, 38(1), 237–257.
- Huh, M.-H. (2001). Translation strategy in EFL writing. *English Teaching*, 56, 75–92.
- Inderawati, R., Hayati, R., Marlina, R., Novarita, N., Awalludin, A., & Anam, S. (2023). Argumentative Essay and Vocabulary Enrichment of English Students by Utilizing Google Translate. *English Community Journal*, 6(2), 131–141.
- JASP Team. (2024). JASP (Version 0.18.3)[Computer software]. <https://jasp-stats.org/>
- Karjo, C. H., & Andreani, W. (2018). Learning Foreign Languages With Duolingo and Memrise. *Proceedings of the 2018 International Conference on Distance Education and Learning*, 109–112.
- Kessler, M., Loewen, S., & Gönülal, T. (2023). Mobile-assisted language learning with Babbel and Duolingo: comparing L2 learning gains and user experience. *Computer Assisted Language Learning*, 1–25.
- Kim, H.-S., Cha, Y., & Kim, N. Y. (2021). Effects of AI chatbots on EFL students' communication skills. *Korean Journal of English Language and Linguistics*, 21, 712–734.
- Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL*, 31(3), 293–311.
- Makino, S., & Tsutsui, M. (1986). *A Dictionary of Basic Japanese Grammar*. Japan Times.
- Makino, S., & Tsutsui, M. (1995). *A Dictionary of Intermediate Japanese Grammar*. Japan Times.
- Makino, S., & Tsutsui, M. (2008). *A Dictionary of Advanced Japanese Grammar*. Japan Times.
- Metruk, R. (2021). The Use of Smartphone English Language Learning Apps in the Process of Learning English: Slovak EFL Students' Perspectives. *Sustainability: Science Practice and Policy*, 13(15), 8205.
- Murtisari, E. T., Bonar, G. J., & Widiningrum, R. (2020). Learning grammar through learning to translate: a means and an end. *The Journal of Asia TEFL*, 17(2), 715–723.
- Rebolledo Font de la Vall, R., & Araya, F. G. (2023). Exploring the benefits and challenges of AI-language learning tools. *The International Journal of Social Sciences and Humanities Invention*, 10(1), 7569–7576.
- Ren, R., Castro, J. W., Acuña, S. T., & de Lara, J. (2019). Evaluation Techniques for Chatbot Usability: A Systematic Mapping Study. *International Journal of Software Engineering and Knowledge Engineering*, 29(11n12), 1673–1702.
- Saleh, N. S., Murtaza, S. F., & Baki, N. U. (2021). A Comparative Analysis on Cohesive Markers in Essay Composition of First Language and Second Language: Direct Writing Versus Translation. *International Journal of Linguistics Studies*, 1(2), 1–10.
- Tavakoli, M., Ghadiri, M., & Zabihi, R. (2014). Direct versus translated writing: The effect of translation on learners' second language writing ability. *GEMA Online Journal of Language Studies*, 14(02), 61–74.
- Tekwa, K. (2024). *Machine Translation and Foreign Language Learning*. Springer Nature.
- Watkins, R. (2023). Guidance for researchers and peer-reviewers on the ethical use of Large Language Models (LLMs) in scientific research workflows. *AI and Ethics*.
- Woollaston, S., Flanagan, B., & Ogata, H. (2024). Chatbots and EFL learning: A systematic review. *Joint Proceedings of LAK 2024 Workshops*, 89–98.
- Wu, X., & Gao, Y. (2011). Applying the extended technology acceptance model to the use of clickers in student learning: Some evidence from macroeconomics classes. *American Journal of Business Education*, 4, 43–50.
- Yang, J. (2022). Perceptions of preservice teachers on AI chatbots in English education. *International Journal of Internet, Broadcasting and Communication*, 14(1), 44–52.
- Zhai, C. (2023). A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. *Computers and Education: Artificial Intelligence*, 4, 100134.
- Zhang, R., & Zou, D. (2022). Types, purposes, and effectiveness of state-of-the-art technologies for second and foreign language learning. *Computer Assisted Language Learning*, 35(4), 696–742.