The JALT CALL Journal
ISSN 1832-4215, 22, 102993 (2026)
https://doi.org/10.29140/jaltcall.2026.102993

Castledown

French learners' use of ChatGPT for interactive textual revision: an in-depth multiple case study

Taegan Holmes
Université Lumière Lyon 2/Laboratoire ICAR, France
taeganholmes@gmail.com

Nicolas Guichon
Université Lumière Lyon 2/Laboratoire ICAR, France
nicolas.guichon@univ-lyon2.fr

Nowadays, generative artificial intelligence (GAI) tools can be found virtually everywhere, including in the second language classroom. Despite this widespread use, there remains a paucity of research exploring how such tools can be utilized for textual revision, especially in languages other than English. Furthermore, the need for students and teachers alike to understand the tools that they use grows ever more important. In the present study, we investigate how French as a Foreign Language students interact with ChatGPT, a widely accessible and well-known GAI tool, to correct their texts. Participants (n = 14) first answered a survey regarding their familiarity with ChatGPT and then participated in a 60-minute workshop on the use of ChatGPT for interactive textual revision in French. Participants (n = 8) wrote a short, narrative text which they then revised using ChatGPT and subsequently refined into a final version. We collected the screen recordings of the writing and revision process, the links to discussion threads and the original and final versions of the texts. We analysed the prompt patterns that emerged, as well as the quality of ChatGPT's responses, revisions and explanations. A sample of participants (n = 2) participated in self-confrontation interviews regarding their experience. Overall, texts saw a significant improvement from first to final draft following interactive revision with ChatGPT regarding error count, though shortcomings concerning ChatGPT's responses as well as the lack of engagement on behalf of participants call into question the pedagogical value of using ChatGPT for textual revision in French.

Keywords: ChatGPT, textual revision, interaction

Introduction

Generative artificial intelligence (GAI or GenAI) seems to have infiltrated virtually every aspect of modern daily life. According to an article in *Harvard Business Review* summarizing a report on how real people are using GAI, over 100 different use-cases were identified which spanned across a wide variety of areas (Zao-Sanders, 2024). From planning workouts to drafting emails to fixing bugs in code to games of Dungeons & Dragons, GAI tools are being widely applied to diverse areas of life. The world of language education is no exception (c.f. Nassau & Molle, 2024).

Among the numerous GAI tools, ChatGPT is the one that seems to have taken precedence, largely due to its ease of use and accessibility. OpenAI launched ChatGPT, a conversational and interactive chatbot, on November 30th, 2022. Within just two months, it boasted a userbase of 100 million active users (Hu, 2023; Kern, 2024), rendering it "[…] the fastest-growing user application in history" (Lo, 2023, p. 1) at that time. ChatGPT, built upon the transformer architecture, belongs to the family of large language models (LLMs) which are trained using immense quantities of data to conversationally respond to user messages (called prompts) in natural language. In language education, it has been posited that ChatGPT has great potential, notably for textual revision and for providing learners with personalized and immediate feedback on their writing (Barrot, 2023; Holmes, 2024a; Lo, 2023; Shi & Aryadoust, 2024; Su et al., 2023).

Research into the use of GAI tools such as ChatGPT in language education is a burgeoning domain with an overwhelming majority of research focusing on English as a Second or Foreign Language (Kern, 2024; Law, 2024). The aim of this in-depth multiple case study is therefore to delve into the didactic potential of ChatGPT in the context of French as a Second or Foreign Language (FSL/FFL) education, a domain in which there is currently a paucity of research (Holmes, 2024b).

In the following sections, we propose a brief literature review, followed by the theoretical underpinnings supporting this study. The research context and methodology are then explained, followed by the results. The data obtained are interpreted and discussed in the following section and in the conclusion, we present the limits of the study and highlight possibilities for future research and next steps.

Literature review

This section aims to provide an overview of the research that has been conducted thus far into the use of ChatGPT for feedback and textual revision to underscore its relevant affordances and limitations identified in the literature.

ChatGPT as a source of feedback

In the context of language education, one of the primary affordances of ChatGPT is its provision of instantaneous, personalized and interactive feedback and revisions (Barrot, 2023; Holmes, 2024a; Lo, 2023; Shi & Aryadoust, 2024; Su et al., 2023). Feedback in second language teaching and learning has

amassed much interest and has been studied using different optics over the years including its degree of directness, i.e. how explicitly it targets an error (Ellis, 2008), the timing of delivery (instantaneous or delayed) (Shintani & Aubrey, 2016) and its source (teacher, self, peer) (Cao & Zhong, 2023; Cho, 2006), among others. Feedback occupies a central role in second language development and acquisition (Bitchener & Ferris, 2011) and is particularly important for improving the skill of written production (Holmes & Hamel, 2025; Woodworth & Barkaoui, 2020).

While it may appear novel to use ChatGPT for feedback and revision, it is far from the first instance of a digital tool being used to this end (see intelligent tutoring systems like E-Tutor [Heift, 2010], automated writing evaluation systems such as Criterion and PEG [Woodworth & Barkaoui, 2020] and correction software like Grammarly [Koltovskaia, 2020], for example). These tools provide automatic feedback (AF) and require the user to choose when they wish to receive feedback, as well as how many times and on how many different drafts (Wilson & Czik, 2016). Furthermore, traditional systems that provide AF (at least prior to recent advances in GAI) tend to provide general (even vague, read surface-level and overly broad) feedback (Ranalli, 2018).

If it can be argued that previous AF systems granted more agency to the user than traditional sources of feedback, then this agency has only increased with the advent of ChatGPT, which may bear both positive and negative attributes. In using both GAI tools and traditional AF systems, the onus remains on the user to submit the text to the tool in order to solicit feedback. However, with GAI tools such as ChatGPT, the responsibility of the user becomes twofold; to provide a text as well as a prompt. This shifts the role of the user from that of a passive "feedback receiver" to an active "feedback seeker" (Yan, 2024). From one perspective, this could allow for more engagement on behalf of the user and result in a more meaningful learning experience. On the other hand, as ChatGPT's responses are prompt-dependent, if a user is not able to craft a high-quality prompt, that is to say a prompt containing pertinent and accurate information as well as clear instructions, this could potentially lead to a failed interaction in which the user becomes frustrated and demotivated and/or accepts suboptimal (or entirely incorrect) feedback, which in turn could actually worsen the quality of their text (Warschauer et al., 2023).

Due to this shift in the role of the user from feedback receiver to feedback seeker, studies have begun to emerge to investigate how second language learners can make use of ChatGPT to solicit feedback and evaluate its quality. That said, there is currently a dearth of information regarding learners' abilities to solicit GAI-based feedback (Yan, 2024).

Despite its importance, the task of critically evaluating the responses provided by ChatGPT is far from straightforward, especially when a user has had no formal training in this domain. A main reason for the opacity of this task is that ChatGPT tends to generate responses that use an authoritative tone and uses "[…] little or no hedging […]" (Kohnke et al., 2023, p. 9), which researchers posit may lead learners (especially those with lower language proficiencies) to systematically and uncritically accept feedback. Furthermore,

M. A. Peters et al. (2023) describe that "[…] to appear a good interlocutor, [ChatGPT] is skewed towards being uncritically affirmative" (p. 15). This can prove to be problematic in the context of interactive feedback where a language learner can challenge the machine and incite it to agree with an incorrect claim, as demonstrated in (Holmes & Hamel, 2025). Beyond being brought to agree with incorrect claims made by the user, ChatGPT can also outright "hallucinate" and generate responses that contain factually incorrect information, all the while framing it as if it were accurate (Heikkilä, 2023; Yan & Zhang, 2024).

Lastly, since ChatGPT's training data was scraped from the internet, and such data contains biases (including harmful, bigoted and incorrect information), ChatGPT's responses are liable to reflect these biases as well, despite efforts to curb this tendency (i.e. Reinforcement Learning by Human Feedback [RLHF]) (Mennella & Quadros-Mennella, 2024). Aside from potentially containing incorrect or harmful information, the majority of ChatGPT's training data is originally in English and subsequently translated into different languages, meaning the tool may tend towards English language constructions. Furthermore, as ChatGPT is trained on textual data, it may also tend toward constructions that are more frequent (or even exclusively used) in written contexts (Kohnke et al., 2023).

While the responsibility to verify ChatGPT's output remains with the user, it is also possible to impact the quality of its responses through prompting strategies. As demonstrated by Coyne et al. (2023) and by Loem et al. (2023), the information (which encompasses its quality, veracity and pertinence) included in a prompt has a significant impact on a system's output. It is therefore important for users to not only be able to craft high quality prompts, but to critically evaluate responses as well.

Despite the limitations of using ChatGPT for feedback, several affordances have been identified as well. Among the most notable are the degree of adaptability and personalization of ChatGPT's feedback and its potential for multilingual contexts and tasks. Regarding adaptability, Shi & Aryadoust affirm that "[t]he advent of transformer-based generative AI systems like ChatGPT promises more personalized feedback" (2024, p. 204). Several other researchers echo the sentiment that a major strength of ChatGPT is the provision of personalized feedback (Barrot, 2023; Kasneci et al., 2023; Kohnke, 2023; Warschauer et al., 2023; Xiao & Zhi, 2023).

ChatGPT for grammatical error correction

In the context of second language education, the fact that ChatGPT's feedback is readily available and highly adaptable is of little importance if it is not of sufficient quality. Studies have begun to emerge focusing on the ability of ChatGPT (and its various predecessors) to perform grammatical error correction (GEC). That said, Coyne et al. (2023) emphasize that this domain is still underexplored and that more research is necessary. Furthermore, the majority of research focuses on English GEC tasks (Kwon et al., 2023).

GEC is generally regarded as a two-step process consisting of error detection and subsequent error correction. Interestingly, a common theme that has emerged regarding ChatGPT's ability for GEC across different languages is that while the model consistently obtains a high score for error detection, it tends to over-correct and introduce more modifications than necessary (Fang et al., 2023; Wu et al., 2023). Due to their generative nature, LLMs like ChatGPT can generate text to express the same idea in a multitude of ways. According to Wu et al. (2023), this may be the cause of ChatGPT's tendency to over-correct.

Theoretical framework

The literature surveyed above highlights the potential merits of ChatGPT for interactive textual revision tasks, while also bringing to light its possible pitfalls or shortcomings for language learners. Whether for good or bad, the current consensus is that generative AI and LLMs are becoming a permanent fixture in our lives (Kasneci et al., 2023; Law, 2024; Mennella & Quadros-Mennella, 2024; Mizumoto & Eguchi, 2023). As candidly put in M. A. Peters et al., "ChatGPT is here to stay" (2023, p. 6). Interestingly, this same quote appears independently in Kohnke et al. (2023, p. 10), further reinforcing the belief that this technology is becoming (or rather has become) integrated into modern life in a permanent fashion. Per Escalante et al., "[i]dentifying and communicating the ethical and appropriate use of AI is an urgent task for practitioners" (2023, p. 3). Specifically, they call for an understanding of how tools like ChatGPT can be integrated into the writing process "[…] in a way that is acceptable to both students and teachers" (Escalante et al., 2023, p. 5). To accomplish these goals, first, empirical research must be conducted.

The importance for students and teachers alike to understand the affordances and limits of the (digital) tools that they use in order to make effective use of them is becoming ever more relevant and there are different frameworks that have been developed to theorize and operationalize these necessary skills. The present study is oriented by that of metatechnolinguistic competences (Guichon, 2024).

Metatechnolinguistic competences are based on "[t]he interweaving of the techno within the metalinguistic to forge this compound adjective [which] is intended to emphasize the crucial role of technologies, which now mediate a large part of language students' communication and have thus become a determining element of literacy in a first language (L1) and additional languages" (Guichon, 2024, p. 2). These competences ultimately refer to the development of a *critical* and *ethical* understanding of the affordances and limitations of technological tools within the context of language learning and teaching.

When it comes to LLMs like ChatGPT, many questions have been evoked pertaining to academic integrity and plagiarism (M. Peters, 2023), as well as content paternity (M. A. Peters et al., 2023). An additional consideration involves privacy, cybersecurity and the protection of personal data (Kasneci et al., 2023). Recent advancements have also raised concerns regarding the deepening of the digital divide, wherein those already in an advantageous

position move ahead, while those who are technologically "less fortunate" find themselves falling yet further behind (Warschauer et al., 2023). Additionally, more general ethical considerations are also worth bearing in mind. For example, as was briefly mentioned above, RLHF was used to help limit the potential biases present in ChatGPT's responses. This process requires human intervention, with workers having to view and subsequently label harmful and bigoted content to protect the end user from its appearance in ChatGPT's output. The workers who undertook this assignment were paid very little (Kern, 2024; Perrigo, 2024) for the work which was labeled as traumatic in nature (Perrigo, 2024). Furthermore, as was mentioned above, despite the use of RLHF, ChatGPT's output is not devoid of biases. Finally, as a generative model, ChatGPT can have a profound environmental impact due to energy consumption (Kasneci et al., 2023).

It is therefore critical to deepen students' metatechnolinguistic competences, in turn diminishing the risk of blind reliance on tools like ChatGPT, instead making way for a critical, meaningful and ethical exploitation. Language educators must first develop their own metatechnolinguistic competences and to accomplish this, more research must be conducted (Guichon, 2024). This is, in sum, the *raison d'être* of the present study.

Methodology

The following section details the methodology that was undertaken during the data collection and its subsequent analysis. We first provide details regarding the participants and the context, followed by the research design and methods followed. We end this section with a description of approaches used during the data analysis.

Data collection

The participants in the present study were 14 students[1] enrolled in a French as a Foreign Language class at an engineering university in France. All participants were briefed on what the study entailed and provided informed consent. The participants had varied linguistic competencies, though according to the instructor's assessment, they all fell around a level B2 per the framework established by the CEFR (Council of Europe, 2001), making them independent users of the target language. This means that the participants were able to express themselves in French, albeit not in a manner devoid of errors. We felt that participants around the B2 level would be the ideal sample for the present research as they would benefit from the revisions provided by ChatGPT due to their intermediate language level and would also have enough of a linguistic basis to converse with ChatGPT about the revision of their text, asking relevant follow-up questions and even challenging the machine if/when necessary.

The participants first answered a short questionnaire focusing on their degree of familiarity with ChatGPT prior to the present project. The questionnaire

[1]While 14 students participated in the initial questionnaire phase of the study, only 8 took part in the writing/revision phase.

was administered online, in French and was fully anonymous. A translated version of the questionnaire is presented in Appendix A. Next, the first author of this publication presented a training module regarding basic background knowledge of ChatGPT, best practices and common pitfalls, as well as several frameworks and strategies for interacting with ChatGPT. While metatechnolinguistic competences were not explicitly mentioned by name during the presentation, the underlying behaviours which comprise them (critical evaluation of responses, information verification using a variety of sources, active engagement with the digital tools being used, etc.) were discussed. The presentation lasted around an hour and was delivered in French. At the end of the session, participants downloaded OBS Studio[2], a screen recording software which was used in the following session to capture the ways that students used ChatGPT to revise their texts.

The following week, the participants were given their writing assignment. They were tasked with writing a text of 100–150 words in French recounting an amusing anecdote of their arrival in France. We chose a narrative text for this task since we wanted the focus to remain on participants' basic writing proficiency in French, rather than on their abilities to synthesize information (c.f. Strobl et al., 2024) or argue a viewpoint (c.f. Yan, 2024). Participants were instructed to write their texts autonomously and then to proceed to an interactive revision process with ChatGPT. They were instructed to first send their text accompanied by the prompt "Revise this text by making the smallest number of modifications possible"[3]. This prompt was chosen to minimize "[…] the erasure of the personal voice of the writer" (Escalante et al., 2023, p. 12) which is a risk often associated with the use of GAI models for writing tasks. ChatGPT would then generate a response with a revised version of the text. The participants were to read this revision and then to send the prompt "Explain all revisions"[4]. ChatGPT would subsequently respond with a list explaining the revisions that were done. From here, participants were encouraged to engage in conversation with ChatGPT to better understand the proposed revisions. They were also encouraged to use the strategies presented the week before and they had access to the presentation during this task. Finally, participants had to write a final version of their text, integrating only the revisions which they deemed necessary to improve the quality and linguistic accuracy of their text. In this way, if a certain revision was seen as superfluous or even counterproductive, it did not have to be included in the final version of the text. The process of writing and revision lasted in total approximately 90 minutes and was recorded using OBS Studio. We collected both the original and the final versions of the texts, the links to the discussion threads with ChatGPT and the screen recordings.

The final stage of data collection consisted in self-confrontation interviews (Guichon, 2009), in which participants were confronted with their past work and asked to elaborate on it in a semi-structured interview setting. The first

[2]https://obsproject.com/
[3]"*Révise ce texte en effectuant le plus petit nombre de modifications possible*" in French.
[4]"*Explique toutes les révisions*" in French.

author viewed all the screen recordings and identified key moments as well as questions aiming to better understand the participants' thought processes and perceptions regarding the experience of using ChatGPT for textual revision. A sample of participants (n = 2) agreed to participate in audio recorded self-confrontation interviews. Interviews lasted approximately 20 minutes each and were conducted in French.

The methodology adopted for the data collection and for its subsequent analysis is presented in Figure 1 below. The dotted lines connecting the different phases of the analysis represent data triangulation, which enabled us to draw better founded and more thorough conclusions. The methods of analysis used are detailed in the following subsection.

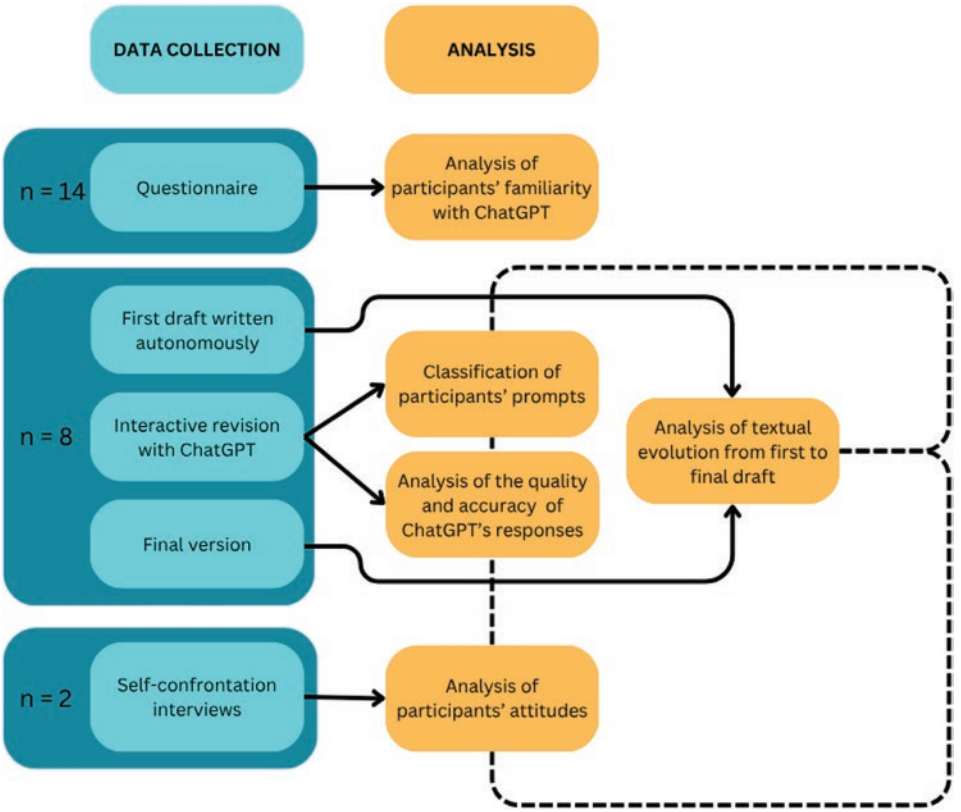

Figure 1. Research methodology and analysis

Methods of analysis

Different methods were undertaken to analyse the various forms of data that were collected. The questionnaire collected quantitative data through closed response questions and one short answer question regarding the participants' first languages. To analyse this data, it was sufficient to count the number of different answers to each question.

In terms of discussion threads, participants' prompts were analysed using an adapted version of the coding scheme created by Yan (2024) on the basis

of prompt content. ChatGPT's responses were evaluated to assess their quality by verifying the accuracy of information contained within each response. We paid particular attention to ChatGPT's responses to the second required prompt in which it lists and explains the corrections. For these responses, we conducted a more granular analysis, again, focusing on the accuracy of their contents. The original and revised versions[5] of the texts were also cross-referenced with the lists of explained modifications allowing us to see if all the corrections in the list were in the revised text and if all the corrections in the revised text were included in the list.

To evaluate the quality of ChatGPT's proposed revisions in response to the first required prompt, we used an adapted version of the framework put forward by Wu et al. (2023) who in turn adapted the methods of Wang et al. (2022), to manually evaluate the revised texts using the metrics of under-corrections, over-corrections and mis-corrections. This allowed for us to simultaneously evaluate ChatGPT's error detection and error correction, the two pillars of GEC. We analysed the corrections generated by ChatGPT on a case-by-case basis, following the methodology proposed by Pfau et al. (2023).

We compared the original versions of participants' texts to the texts that they submitted as the final version on the basis of error count to evaluate textual evolution and improvement. Since defining an error is a thorny and subjective task, we opted to instead count the number of modifications necessary to rectify the errors[6].

Lastly, the self-confrontation interviews were transcribed and subsequently analysed using thematic analysis (Braun & Clarke, 2006). The interviews were manually transcribed by the first author, though two different automatic speech-to-text software models were tested as well. These models proved to be unsatisfactory, producing inaccurate transcriptions requiring extensive corrections, so the manual transcription was the one used for the analysis.

Results

In the following section, we present the results of our study. We begin with the responses to the questionnaire, followed by the data pertaining to the discussion threads, including the classification of participants' prompts and the quality analysis of ChatGPT's responses. Next, we present the results of the analysis of ChatGPT's revised version of the text in response to the first required prompt, as well as an in-depth look at ChatGPT's revision explanations. An evaluation of participants' final versions of their texts is then presented and finally, we cover the themes which emerged during the self-confrontation interviews. To more clearly illustrate the task that was carried out, we decided to include an excerpt of a participant's original text. Table 1 depicts sample sentences from

[5]The initial revision generated as a response to the first required prompt.

[6]For example, the sentence "**J'ai pas vu*" could be seen as containing only one error, a negation error. That said, to correct it and obtain the correct form *"Je n'ai pas vu"*, it would be necessary to implement two modifications, one which modifies the form of the subject and one which adds the missing negative particle. One could argue that it therefore contains two errors. In the present article, the number of errors refers to the number of modifications necessary to render the text correct.

Table 1. Excerpt of Hugo's initial text with English translation

Original [French]	Translation [English]
L'alarme d'incendies a été activé plusiers fois pendant l'année mais c'était toujours "par accident". Après deux mois, on m'a dit que cet individu qui n'habite pas dans le bâtiment active de manière déliverée l'alarme pour faire rigoler ses copains.	The fires alarm was activated meny times during the year but it was always "by accident". After two months, I was told that this individual who does not live in the building deliverately activates the alarm to make his friends laugh.

the original text written by Hugo[7], as well as an English translation proposed by the first author. The errors in the English translation aim to mirror those in the original version.

Questionnaire responses

The goal of the questionnaire that was administered at the beginning of the present study aimed to gather information regarding participants' familiarity with ChatGPT. Aside from the first question, which was a short answer response regarding the languages that participants spoke at home (*i.e.* mother tongues), the questionnaire was comprised of nine other multiple-choice questions regarding habits surrounding the use of ChatGPT. Table 2 presents the distribution of participants' first languages per their responses to the first question.

Table 2. Distribution of participants' first languages

First languages	Number of participants
Arabic	2
English	2
Finnish	1
German	2
Romanian	1
Spanish	3
Turkish	2
Ukrainian	1

All of our participants (100%[8]) responded that they already had ChatGPT accounts prior to their involvement in the present study. The majority of participants (57%) created their accounts in 2023 with remaining participants signing up for ChatGPT in 2022 (21%), in 2024 (14%) and in 2025 (7%). Almost all participants (86%) use the free version of ChatGPT, with only 14% using the paid version. Regarding uses of ChatGPT, two participants (14%)

[7]Participants were given pseudonyms.
[8]All percentages are rounded to the nearest integer.

responded that they used the tool for learning French. Only one participant (7%) responded that they used ChatGPT for learning other languages. All but one participant (93%) responded that they used ChatGPT for tasks other than language learning.

Of the two participants who indicated that they used ChatGPT for learning French, one declared using it several times a month, while the other used it less than once a month. The participant who used ChatGPT for learning another language did so several times per month. Of the thirteen participants who used ChatGPT for tasks other than language learning, five (38%) do so daily, six (46%) use it several times a week, one participant (8%) uses it once a week and one participant (8%) uses it less than once a month.

Discussion threads

Despite receiving fourteen responses to the initial questionnaire, only eight participants completed the subsequent writing and revision task with ChatGPT, due to a low attendance rate and technical problems. The corpus of discussion threads is therefore comprised of only eight. In total, the corpus of eight discussion threads contains 108 messages (54 prompts sent by the participants and 54 responses from ChatGPT). On average, discussion threads are comprised of 13.5 messages. The longest discussion thread, belonging to Basil, contains 34 messages (17 prompts and 17 responses) and the shortest, that of Derek, contains only four messages (two prompts and two responses). All participants used at minimum the two required prompts and all but one participant (Derek) delved deeper into the revision process, making their own prompts, namely, to gain a better understanding of the provided revisions or to further improve their texts. As mentioned in the methods section, we dissected the discussion threads and analysed the prompts using an adapted version of the coding scheme created by Yan (2024) and we also evaluated the quality of ChatGPT's responses based on accuracy of information.

Typology of prompts

In analysing the prompts in the discussion threads, we sought to investigate the types of requests that participants were making, as well as how they responded to ChatGPT's answers. The typology that we adopted was based on that of Yan (2024), which ascribed different tags to the prompts based on their content. We used the typology proposed by Yan (2024) as a starting point, but we also created new categories which fit our data. Therefore, we adopted an abductive approach and used both deductive and inductive coding. Table 3 presents the different tags that were used, their descriptions, as well as the number of times they occurred.

As is illustrated in Table 3, the participants used mostly bare prompts, which aimed to obtain information (feedback or otherwise) through an instruction or a question. The tag with the second highest number of occurrences was text sharing, designating messages wherein any version of the text was sent. There was a total of 18 instances of messages containing one of the required prompts. Specifications were identified eight times and included instances where an

Table 3. Participant prompt descriptors and occurrences inspired by Yan (2024)

Tag	Description	Number of occurrences
Bare prompt [BP]	A simplistic prompt containing either an instruction or a question (such as asking for the word count).	28
Text share [TS]	The sending of any version of the text.	20
Bare prompt – Required [BPR]	One of the required prompts (either requesting a revision with minimal edits or the explanation of those edits).	18
Specification [SPE]	An addition regarding a specific aspect (for example, specifying that the rewritten text should convey that the subject of the text is present in the room where it is being read).	8
Desired response [DR]	A specification describing the format of the desired response (such as text length).	5
Tone [TON]	A specification regarding the tone/style of the text (such as requesting a more formal version).	4
Background [BG]	A provision of background or of contextual information (for example, the participant's first language).	3
Persona [PER]	An imposition of a persona (for example, telling ChatGPT to adopt the role of a French teacher).	2
Acceptation [ACC]	An acceptation of something proposed in a previous ChatGPT message.	1
Affective [AFF]	An affective/attitudinal evaluation.	1
Example [SHOT]	The provision of an example, as seen in the "Few-shot" prompting context.[9] This prompting strategy was shown in the presentation.	1
Quality [QUA]	An evaluation of quality (for example, specifying that a previous answer was not of sufficient quality).	1

additional requirement was tacked on to a bare prompt, for example, when Hugo specified in a message that the text needed to be rewritten "to make the audience laugh". There were five occasions where the desired output was specified, such as desired length or format of responses. We tagged instances that focused on tone/style four times across the corpus and instances providing background or contextual information only thrice. We identified two occasions where a participant told ChatGPT to adopt a role/persona. Finally, for the last four tags, one instance each was identified; a proposal from ChatGPT for it to improve the flow of the text was accepted, there was an appreciation of the revised text, an example was explicitly shared in an effort to improve response accuracy, and there was one evaluation of the quality of provided feedback.

Evaluation of ChatGPT's responses

The following step in the analysis involved evaluating the quality of the responses generated by ChatGPT. We looked first at the overall proportion of incorrect responses provided by ChatGPT compared to the total number of

[9]See Kojima et al. (2023).

responses it generated. We then analysed the quality of ChatGPT's response to the first required prompt, that is, the initial revision of the text. Finally, we examined ChatGPT's responses to the second required prompt, the explanations of all the revisions that it carried out during the initial revision.

Quality of responses

We examined all of ChatGPT's responses to see the amount containing errors and misinformation. Out of the total 54 responses, 37 contained at least one error or point of misinformation, which equates to 69%. These inaccuracies included, for example, incorrect word counts (from the discussion threads of Adam and Basil) which were off by as much as 12%, as well as vocabulary mistakes, as was seen with Hugo, where ChatGPT claims to have generated an *accusative* sentence, when in fact it generated an *accusatory* sentence.

Quality of the initial revision

As mentioned above, in analysing the quality of the initial revision, we adopted the framework used by Wu et al. (2023) to account for under-corrections, over-corrections and mis-corrections. Under-corrections refer to cases where a correction was needed, but was not done, therefore leaving an erroneous form untouched. Over-corrections are superfluous modifications and refer to any extra changes that are done without being strictly necessary. Lastly, mis-corrections are occasions where a modification is required, though an incorrect one is made. This includes modifications that address one error while introducing another. These metrics were applied using human evaluation conducted by the first author. Table 4 contains the results of this analysis for each participant, as well as a total count.

Table 4. Human evaluation of erroneous corrections

Participant	Under-corrections	Over-corrections	Mis-corrections
Adam	3	2	0
Basil	1	4	0
Christina	0	25	0
Derek	1	0	0
Emma	0	4	2
Flavia	1	3	1
Gary	1	8	0
Hugo	1	7	0
Total	8	53	3

As can be seen in Table 4, ChatGPT's revisions rarely missed necessary modifications, averaging 1 (standard deviation of 0.926) under-correction per initial revision. Even fewer mis-corrections were identified, with an average of

less than one (0.375 with a standard deviation of 0.744) per initial revision. Conversely, it generated many more unnecessary corrections, resulting in an average of 6.625 (standard deviation of 7.855) over-corrections per initial revision.

Quality of the explanations of the initial revision

Previously, we looked at the total amount of ChatGPT's responses which contained errors. Since a principal objective of the present study was also to analyse the use of ChatGPT to revise text in French, we decided to take a closer look at its responses containing the explanations of all revisions. To do this, we looked exclusively at the responses containing the explanation of revisions and we verified every point of information and noted each individual error as well.

We looked at the initial revisions to see how many modifications were generated by ChatGPT and cross-referenced this information with the number of modifications listed in its explanation of all the revisions. This allowed us to see if ChatGPT's explanations of all modifications were in fact comprehensive. We also checked if ChatGPT "invented" any errors in its list of explanations (*i.e.* errors which were not present in the original text submitted by a participant, but that ChatGPT claimed to have found). Similarly, we looked for occasions where ChatGPT listed a modification in its explanation of revisions, despite said modification being absent from the revised version. Lastly, we evaluated the amount of incorrect information like inaccurate descriptions of modifications (see Figure 2) and inconsistencies in the points contained within the explanation, such as instances where ChatGPT arbitrarily modified a verb tense in one sentence and not in another. Table 5 contains this information.

2. "premiere" → "première"

→ Ajout de l'**accent grave** sur le premier "e", car "première" est au féminin pour accorder avec "fois".

Figure 2. Incorrect explanation of a modification in Adam's text

Table 5. Evaluation of the quality of ChatGPT's explanations

	Adam	Basil	Christina	Derek	Emma	Flavia	Gary	Hugo	Total
Initial Revision Modifications	49	23	62	34	39	34	47	31	319
Modifications in the Explanation	34	17	56	27	31	26	37	22	250
Invented Errors	0	0	0	1	0	1	0	0	2
Invented Modifications	1	1	0	0	0	0	0	0	2
Points of Erroneous Information	8	4	7	0	5	3	6	2	35
Points of Inconsistency	1	1	2	0	2	2	1	1	10

While a total of 319 modifications were implemented in the initial revisions of participants' texts, only 250 of these were mentioned in the explanations of all the modifications carried out. ChatGPT therefore listed 78% of the modifications made, despite being explicitly prompted to explain *all* revisions. We also identified two instances where ChatGPT invented errors, claiming that they were in the original texts when in fact, they were not. There were two instances where ChatGPT responded with a claim of having made modifications, despite said modifications not being found in its revision. We found 35 points of incorrect information or committed errors within ChatGPT's responses explaining its revisions. For example, when explaining the revisions carried out in Adam's text, ChatGPT's answer claimed that to correct the erroneous form "**premiere*", a grave accent was needed on the first "e". Despite this inaccurate explanation, the error was properly addressed (with the form "*première*", putting the grave accent on the second "e") in both the revised text and in the explanation; the necessary correction was simply incorrectly explained (see Figure 2).

Another example of incorrect information in ChatGPT's explanations of revisions carried out comes from Emma's discussion thread (Figure 3). In ChatGPT's response, it was rightly identified that the pronoun "*un*" had to be replaced by the pronoun "*une*", though rather than correctly explaining the reason behind this replacement (the pronoun was referring to the antecedent "*personne*", which, in French is a feminine noun), ChatGPT's response claimed that this change was necessary because the feminine gender was more likely in the administrative context in which the anecdote took place. This explanation is not only incorrect, but also sexist and highly problematic.

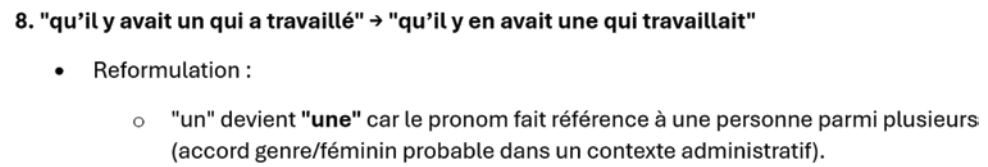

Figure 3. Excerpt of an incorrect explanation of a modification in Emma's text

Lastly, we identified 10 inconsistencies/incoherences in ChatGPT's revision explanations, including instances where it was claimed that a type of revision was necessary despite not implementing it elsewhere (for example, in its explanation of revisions done for Adam's text where it arbitrarily switched between different forms of the past tense). It also sometimes provided explanations out of order (which occurred twice in its explanation of revisions for Christina's text).

Comparison of participants' original texts and final versions

We compared each original text with what participants submitted as their final versions on the basis of overall quantity of errors. Unsurprisingly, the final versions always contained significantly less errors than the original versions. Participants' original texts each required on average 30 modifications (standard deviation of 8.418), whereas final versions were occasionally devoid of errors, requiring no further modifications. These data are presented in Table 6.

Table 6. Comparison of error counts in original and final versions

Participant	Errors in original text	Errors in final version
Adam	44	2
Basil	20	0
Christina	23	0
Derek	35	1
Emma	34	0
Flavia	25	7
Gary	36	1
Hugo	23	4

Participants' attitudes and perceptions

The final step in our analysis consisted in examining participants' attitudes and perceptions of using ChatGPT for interactive textual revision that were evoked during the semi-structured interviews. Two participants (Flavia and Hugo) were interviewed. The transcripts of the two interviews were analysed using thematic analysis (Braun & Clarke, 2006). We identified four major themes across the two interviews: overall experience/appreciation, perceived affordances, perceived limitations and content paternity. Due to the small sample size, it would be presumptuous to attempt to generalize our findings from this stage of the analysis. Still, we find it relevant to briefly mention the more salient points that were evoked.

In terms of overall appreciation, both participants that were interviewed said that they enjoyed the experience of using ChatGPT for revising their texts. Interestingly, both participants also had a high level of trust in ChatGPT within the context of the study, in part due to perceived low proficiency in French. When it came to affordances, both Flavia and Hugo cited ChatGPT's usefulness for spelling corrections and expressed an appreciation for its interactive nature which allowed them to ask for further explanations. Flavia specified that the best part of using ChatGPT for textual revision was being able to ask why certain corrections were done.[10] As for limitations, Flavia found it tedious to manually implement ChatGPT's suggested corrections and Hugo mentioned that ChatGPT generated too much text.

Both participants had opposing takes on content paternity and feelings of ownership over their final texts. This is understandable, especially since they each took very different approaches to producing said texts. Flavia felt ownership over the final text and had no reservations about submitting it for a hypothetical class assignment, remaining steadfast in the opinion that ChatGPT did not write the text. To produce the final version of the text, Flavia copied the

[10]Excerpt from Flavia's interview in French : 02:48 "*Je crois que le meilleur, c'est de demander pourquoi il a fait les corrections […]". In English, this excerpt reads : "I think that the best, it is asking why it did the corrections […]".

initial version into a Word document and individually applied the manual corrections suggested by ChatGPT, albeit, overlooking some. This explains the elevated error count in Flavia's final text.

In contrast, Hugo felt no ownership over the final text, feeling incapable of producing it alone and did not feel comfortable with the idea of submitting it as a class assignment. To produce the final version of the text, Hugo copied a version of the text that was rewritten by ChatGPT, almost in its entirety. The only part left out was the final sentence, which was replaced with a different closing sentence proposed by ChatGPT in a later message.

Discussion

In this penultimate section, we aim to explicate the results of the present study in relation to the literature. To maintain textual cohesion, we address our results in the same order in which they are presented in the previous section.

The questionnaire that was administered prior to the writing and interactive revision task aimed to gather information regarding the participants' knowledge of ChatGPT. All participants stated having ChatGPT accounts prior to the study and the majority responded that they used it for tasks other than language learning. The widespread use of ChatGPT by students is unsurprising, as other studies have identified an elevated usage rate among student populations (c.f. Hellmich, 2024).

The eight discussion threads that were collected and subsequently analysed revealed interesting trends in terms of prompt engineering. As was the case in Yan (2024), the tag that was most used to categorize our participants' prompts was BP (for bare prompt), meaning participants often constructed a minimalistic prompt. Participants in both studies also tended to include specific requests in their prompts, for example, asking for a specific tone to be used or specifying the length of response that was expected.

The vast majority of ChatGPT's answers (69%) contained at minimum one error/point of misinformation. While concerning, this proportion is not surprising, as other researchers have identified elevated error counts and instances of inaccuracies in ChatGPT's responses within the specific context of textual revision and feedback provision (Alsofyani & Barzanji, 2024; Arifin et al., 2024; Bai & Wei, 2024; Guo, 2024; Poláková & Ivenz, 2024; Tam, 2024).

In terms of revision quality, we found that in the initial revisions that ChatGPT generated, errors were seldom overlooked and even more rarely mis-corrected, however, ChatGPT generated a high number of over-corrections, despite being instructed to perform the smallest number of modifications possible. These results are consistent with the findings of Fang et al. (2023), Coyne et al. (2023) and Wu et al. (2023), which found that different versions of ChatGPT, though adept at identifying issues within writing, tended to over-correct, meaning they introduced superfluous and unnecessary modifications to already correct segments.

In terms of the quality of ChatGPT's explanations of revisions, we identified several issues. Firstly, ChatGPT never responded with a list of all the modifications that it made, despite being explicitly asked to provide an exhaustive

list. We identified 35 points of blatantly incorrect information or errors in ChatGPT's explanations of modifications, as well as 10 points of confusion/inconsistency. These errors evoke the notion of hallucinations (Heikkilä, 2023; Yan & Zhang, 2024) and the fact that LLMs like ChatGPT "[…] hold no interpretation of truth" (Warschauer et al., 2023, p. 5).

As a generative model, ChatGPT cannot reason or think, instead, it relies on its training data to generate answers. Since the vast majority of this data was scraped from the internet, the risk of it containing inaccurate and biased information is high. One of ChatGPT's more striking errors was when a response claimed that a properly identified error (the masculine gender being used when the female was necessary) had to be corrected not for a legitimate grammatical reason, but instead because it was more likely that the pronoun in question was referring to a woman, due to the administrative context in which the anecdote recounted took place. This underlines the point brought up by Warschauer et al. (2023) that ChatGPT can produce responses containing harmful biases.

Despite the issues identified in the discussion threads, participants' texts did undergo significant improvement in terms of error count, meaning final versions systematically contained less errors than their original counterparts. However, this was found to be less a factor of participants implementing individual, manual corrections, but in fact because most participants submitted a version of their text reworked by ChatGPT. While in terms of number of errors, texts saw significant improvement from the initial version to the final one, this brings into question the authorship of the text and consequently, issues of plagiarism and academic integrity (M. Peters, 2023; M. A. Peters et al., 2023).

There is also the issue of the pedagogical value of using ChatGPT as a tool for writing in a second language. The way in which most participants produced the final version of their texts is a potential cause for concern. Rather than utilizing ChatGPT as a tool for revision and to model proper structures in the target language, the majority of participants uncritically adopted all of ChatGPT's proposed feedback, submitting a version of their text rewritten by ChatGPT as their final text. This harkens back to the issue of engagement in the revision process as well as the critical evaluation of ChatGPT's output (Kohnke et al., 2023) and technological overreliance (Guichon, 2024). If participants are content to submit a text produced by ChatGPT (albeit sourced from their original text) without actively engaging in the revision process, this calls into question the potential pedagogical value of this application. In our study, since the majority of participants did not demonstrate a critical/meaningful use of ChatGPT during the revision process, then they did not significantly (if at all) develop their metatechnolinguistic competences (Guichon, 2024). Participants did not seem overly committed to a critical, meaningful and ethical exploitation of the tool, opting instead for blind reliance and uncritical acceptations of ChatGPT's feedback, even when it was unnecessary or even incorrect. Through our analysis of the screen recordings, the discussion threads and the different versions of participants' texts, we observed a lack of critical engagement with ChatGPT not only during the interactions comprising the revision process, but

also during the composition of the final versions. This was demonstrated when participants did not attempt to counter ChatGPT when it provided inconsistent or even incorrect responses. Moreover, the fact that a majority of participants produced their final texts by copying and pasting ChatGPT's revisions further exemplifies this lack of critical engagement and, by extension, the development of metatechnolinguistic competences.

The two semi-structured interviews that were conducted revealed interesting information regarding how the participants perceived using ChatGPT for revising their texts. Both participants had a positive perception of the overall experience which aligns with the findings of Holmes & Hamel (2025). The participants also each identified salient affordances and limitations. Both participants had essentially opposing approaches to creating their final copies and subsequently, also viewed the question of authorship in very different ways. Flavia manually implemented the suggested revisions from ChatGPT and felt ownership over the final product. In contrast, Hugo's final text was a version rewritten by ChatGPT, modified to include another suggestion from ChatGPT as the closing sentence. Hugo did not feel ownership over this text. This difference in feelings of ownership may be attributed to the ways in which the final texts were produced. Flavia was actively engaged with the text while producing her final version and therefore retained a feeling of ownership over it. Conversely, Hugo was minimally involved with the production of his final text, resulting in him feeling the text was no longer his own.

The combination of ChatGPT's high ratio of errors in its responses, its significant proportion of over-corrections during textual revision, the inaccuracies within its explanations of said revisions, aforementioned ethical concerns as well as the uncritical manner in which most participants used the tool does not make for a very positive outlook regarding ways in which ChatGPT can be integrated into FSL/FFL textual revision. The results of the present study further underline the imperativeness of making language learners aware of the limitations of tools like ChatGPT if they are to be used during the revision process (Guichon, 2024; Strobl et al., 2024).

Conclusion

At the inception of this project, our thinking was that if ChatGPT could generate a high-quality revision as well as satisfactory explanations behind its modifications, the tool could prove useful for textual revision in the context of FSL/FFL learning. While the original revisions provided by ChatGPT were generally of decent quality (though did not adhere to the minimal edits paradigm that was put in place), its explanations of revisions contained a disconcerting number of errors. Worse still was its overall quantity of incorrect responses. These results call into question the usability of ChatGPT as a tool for textual revision in French, not only for the shortcomings of the model itself, but for the disengaged manner in which most participants used it to revise their texts. We therefore feel that if ChatGPT is to be used in the context of textual revision by French learners, it is necessary for them to understand the tool's limitations regarding not only linguistic inaccuracies, but also harmful biases that it may

propagate in its answers. Specific examples (such as those cited in this article) may be used to help learners better comprehend these limitations. Furthermore, it is important for learners to understand the difference between using ChatGPT to revise and using ChatGPT to write. We therefore recommend that clear policies and definitions be established and discussed with learners to optimize their use of not only ChatGPT, but other similar digital tools.

Before concluding the article, we would like to address limitations pertaining to the present study, as well as potential areas for future research. We found a primary limitation to be our sample size, which decreased further for the actual writing and revision task, though, being a case study, this was to be expected. Our study also spanned only two weeks, with participants attending the information session during the first week and participating in the hands-on experience the next. Finally, with regards to our analysis of ChatGPT's revisions, Mizumoto et al. (2024) identified potential issues with the methods used by Pfau et al. (2024), which we in turn adopted to evaluate the quality of ChatGPT's revisions in our study. Mizumoto et al. argue that using the system's output as the initial standard for error identification "[…] inadvertently establishes ChatGPT as the benchmark for correctness […]" (2024, p. 3), though, as we stated in our methodology, this was the most practical approach.

In future research, it would be beneficial to have a larger sample size in order to further validate and generalize the results. A study involving more participants is currently underway. It could also be interesting to prohibit participants from submitting a text fully generated/rewritten by ChatGPT, requiring them instead to systematically critically evaluate each proposed revision before accepting it and integrating it into their rewritten text, or rejecting it and excluding it.

Though the results of the present study do not seem to promote the use of ChatGPT as a tool during the revision process in FSL/FFL learning, they did bring about some interesting findings. Said findings could be used to further improve ChatGPT, for example, by training it to better adhere to the minimal edits paradigm during textual revision. Ultimately, these results can serve as a springboard for future studies and/or be used to bring awareness to the shortcomings and risks of using ChatGPT in the context of textual revision.

References

Alsofyani, A. H., & Barzanji, A. M. (2024). The effects of ChatGPT-generated feedback on Saudi EFL learners' writing skills and perception at the tertiary level: A mixed-methods study. *Journal of Educational Computing Research, 63*(2), 431–463. https://doi.org/10.1177/07356331241307297

Arifin, M. A., Rahman, A. A., Balla, A., Susanto, A. K., & Pratiwi, A. C. (2024). ChatGPT affordances and Indonesian EFL students' perceptions in L2 writing: A collaborative reflexive thematic analysis. *Changing English*, 1–17. https://doi.org/10.1080/1358684X.2024.2418132

Bai, L., & Wei, Y. (2024). Exploring EFL learners' integration and perceptions of ChatGPT's text revisions: A three-stage writing task study. *IEEE Transactions*

on Learning Technologies, *17*, 2215–2226. https://doi.org/10.1109/TLT.2024.3491864

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, *57*. https://doi.org/10.1016/j.asw.2023.100745

Bitchener, J., & Ferris, D. (2011). *Written corrective feedback in second language acquisition and writing*. Routledge. https://doi.org/10.4324/9780203832400

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cao, S., & Zhong, L. (2023). *Exploring the effectiveness of ChatGPT-based feedback compared with teacher feedback and self-feedback: Evidence from Chinese to English translation*. ArXiv. https://doi.org/10.48550/arXiv.2309.01645

Cho, K., Schunn, C. D., & Charney, D. (2006). Commenting on writing: Typology and perceived helpfulness of comments from novice peer reviewers and subject matter experts. *Written Communication*, *23*(3), 260–294. https://doi.org/10.1177/0741088306289261

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Language Policy Unit, Strasbourg. https://rm.coe.int/1680459f97

Coyne, S., Sakaguchi, K., Galvan-Sosa, D., Zock, M., & Inui, K. (2023). *Analyzing the performance of GPT-3.5 and GPT-4 in grammatical error correction*. ArXiv. http://arxiv.org/abs/2303.14342

Ellis, R. (2008). A typology of written corrective feedback types. *ELT Journal*, *63*(2), 97–107. https://doi.org/10.1093/elt/ccn023

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, *20*. https://doi.org/10.1186/s41239-023-00425-2

Fang, T., Yang, S., Lan, K., Wong, D. F., Hu, J., Chao, L. S., & Zhang, Y. (2023). *Is ChatGPT a Highly Fluent Grammatical Error Correction System? A Comprehensive Evaluation*. ArXiv. http://arxiv.org/abs/2304.01746

Guichon, N. (2009). Training future language teachers to develop online tutors' competence through reflective analysis. *ReCALL*, *21*(2), 166–185. http://dx.doi.org/10.1017/S0958344009000214

Guichon, N. (2024). Sur les traces de Richard Kern: Acknowledging the pivotal role of technologies in language education. *The Modern Language Journal*, *108*(2), 563–566. https://doi.org/10.1111/modl.12931

Guo, X. (2024). Facilitator or thinking inhibitor: understanding the role of ChatGPT-generated written corrective feedback in language learning. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2024.2445177

Heift, T. (2010). Developing an intelligent language tutor. *CALICO Journal*, *27*(3), 443–459. https://www.jstor.org/stable/calicojournal.27.3.443

Heikkilä, M. (2023, April 12). AI literacy might be ChatGPT's biggest lesson for schools. MIT *Technology Review*. https://www.technologyreview.com/2023/04/12/1071397/ai-literacy-might-be-chatgpts-biggest-lesson-for-schools/

Hellmich, E. A., Vinall, K., Brandt, Z. M., Chen, S., & Sparks, M. M. (2024). ChatGPT in language education: Centering learner voices. *Technology in Language Teaching & Learning*, *6*(3). https://doi.org/10.29140/tltl.v6n3.1741

Holmes, T. (2024a). ChatGPT pour la rétroaction corrective écrite interactive en apprentissage du français langue seconde [Master's Thesis, University of Ottawa]. uO Research. https://doi.org/10.20381/ruor-30407

Holmes, T. (2024b). Actes des XXXVII journées de linguistique. Dans G. Frazer-McKee, N. Gignac & L. Wong (Dir.), *L'utilisation de ChatGPT 3.5 pour la rétroaction corrective écrite interactive en enseignement-apprentissage du français langue seconde : une étude exploratoire* (p. 20–35). Université Laval. https://doi.org/10.70637/xvc8m613

Holmes, T., & Hamel, M.-J. (2025). Analyse des interactions d'apprenants de FLS avec ChatGPT pour la rétroaction corrective écrite. *Alsic, 28*(1). https://doi.org/10.4000/13f6h

Hu, K. (2023, February 2). ChatGPT sets record for fastest-growing user base—Analyst note. *Reuters*. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*. https://doi.org/10.1016/j.lindif.2023.102274

Kern, R. (2024). Twenty-first century technologies and language education: Charting a path forward. *The Modern Language Journal*, *108*(2), 515–533. https://doi.org/10.1111/modl.12924

Kohnke, L., Moorhouse, B., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, *54*(2). https://doi.org/10.1177/00336882231162868

Kojima, T., Gu, S. S., Reid, M., Matsuo, Y. & Iwasawa, Y. (2023). *Large Language Models are Zero-Shot Reasoners.* ArXiv. https://doi.org/10.48550/arXiv.2205.11916

Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, *44*. https://doi.org/10.1016/j.asw.2020.100450

Kwon, S. Y., Bhatia, G., Nagoud, E. M. B., & Abdul-Mageed, M. (2023). *ChatGPT for Arabic Grammatical Error Correction*. ArXiv. http://arxiv.org/abs/2308.04492

Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, *6*. https://doi.org/10.1016/j.caeo.2024.100174

Lo, C. K. (2023). What is the impact of ChatGPT on Education? A rapid review of the literature. *Education Sciences*, *13*(4). https://doi.org/10.3390/educsci13040410

Loem, M., Kaneko, M., Takase, S., & Okazaki, N. (2023). *Exploring effectiveness of GPT-3 in grammatical error correction: A study on performance and controllability in prompt-based methods*. ArXiv. http://arxiv.org/abs/2305.18156

Mennella, T., & Quadros-Mennella, P. (2024). Student use, performance and perceptions of ChatGPT on college writing assignments. *Journal of University Teaching and Learning Practice*, *21*(1). https://doi.org/10.53761/pgwk1a93

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, *2*(2). https://doi.org/10.1016/j.rmal.2023.100050

Mizumoto, A., Yasuda, S., & Tamura, Y. (2024). Identifying ChatGPT-generated texts in EFL students' writing: Through comparative analysis of linguistic fingerprints. *Applied Corpus Linguistics*, *4*(3). https://doi.org/10.1016/j.acorp.2024.100106

Nassau, G. & Molle, N. (2024, June 12–15). *Uses of ChatGPT by French as a foreign language students* [Plenary Session] AI and teachers, CercleS Seminar, Nancy, France.

Perrigo, B. (2023, January 18). Exclusive: The $2 per hour workers who made ChatGPT Safer. *TIME*. https://time.com/6247678/openai-chatgpt-kenya-workers/

Peters, M. (2023). Note éditoriale: Intelligence artificielle et intégrité académique peuvent-elles faire bon ménage ? *Revue des sciences de l'éducation*, *49*(1). https://doi.org/10.7202/1107846ar

Peters, M. A., Jackson, L., Papastephanou, M., Jandric, P., Lazaroiu, G., Evers, C. W., Cope, B., Kalantzis, M., Araya, D., Tesar, M., Mika, C., Chen, L., Wang, C., Sturm, S., Rider, S., & Fuller, S. (2023). AI and the future of humanity: ChatGPT-4, philosophy and education - Critical responses. *Educational Philosophy and Theory*, *56*, 828–262. https://doi.org/10.1080/00131857.2023.2213437

Pfau, A., Polio, C., & Xu, Y. (2023). Exploring the potential of ChatGPT in assessing L2 writing accuracy for research purposes. *Research Methods in Applied Linguistics*, *2*(3). https://doi.org/10.1016/j.rmal.2023.100083

Poláková, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education*, *11*(1). https://doi.org/10.1080/2331186X.2024.2410101

Ranalli, J. (2018). Automated written corrective feedback: how well can students make use of it? *Computer Assisted Language Learning 31*(7), 653–674. https://doi.org/10.1080/09588221.2018.1428994

Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL*, *36*(2), 187–209. https://doi.org/10.1017/S0958344023000265

Shintani, N. & Aubrey, S. (2016). The effectiveness of synchronous and asynchronous written corrective feedback on grammatical accuracy in a computer-mediated environment. *The Modern Language Journal*, *100*(1), 296–319. https://doi.org/10.1111/modl.12317

Strobl, C., Menke-Bazhutkina, I., Abel, N., & Michel, M. (2024). Adopting ChatGPT as a writing buddy in the advanced L2 writing class. *Technology in Language Teaching & Learning*, *6*(1). https://doi.org/10.29140/tltl.v6n1.1168

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, *57*. https://doi.org/10.1016/j.asw.2023.100752

Tam, A. C. F. (2024). Interacting with ChatGPT for internal feedback and factors affecting feedback quality. *Assessment & Evaluation in Higher Education, 50*(2), 219–235. https://doi.org/10.1080/02602938.2024.2374485

Wang, W., Jiao, W., Hao, Y., Wang, X., Shi, S., Tu, Z., & Lyu, M. (2022). Understanding and improving sequence-to-sequence pretraining for neural machine translation. In S. Muresan, P. Nakov, & A. Villavicencio (Eds.), *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 2591–2600). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-long.185

Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q. & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *Journal of Second Language Writing, 62*. https://doi.org/10.1016/j.jslw.2023.101071

Wilson, J., & Czik, A. (2016). Automated essay evaluation software in English language arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers & Education, 100*, 94–109. https://doi.org/10.1016/j.compedu.2016.05.004

Woodworth, J., & Barkaoui, K. (2020). Perspectives on using automated writing evaluation systems to provide written corrective feedback in the ESL classroom. *TESL Canada Journal, 37*(2), 234–237. https://doi.org/10.18806/tesl.v37i2.1340

Wu, H., Wang, W., Wan, Y., Jiao, W. & Lyu, M. R. (2023). *ChatGPT or Grammarly? Evaluating ChatGPT on grammatical error correction benchmark.* ArXiv. https://doi.org/10.48550/arXiv.2303.13648

Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages*, *8*(3). https://doi.org/10.3390/languages8030212

Yan, D. (2024). Feedback seeking abilities of L2 writers using ChatGPT: A mixed method multiple case study. *Kybernetes 54*(7). https://doi.org/10.1108/K-09-2023-1933

Yan, D., & Zhang, S. (2024). L2 writer engagement with automated written corrective feedback provided by ChatGPT: A mixed-method multiple case study. *Humanities and Social Sciences Communications*, *11*(1). https://doi.org/10.1057/s41599-024-03543-y

Zao-Sanders, M. (2024, March 19). How people are really using GenAI. *Harvard Business Review*. https://hbr.org/2024/03/how-people-are-really-using-genai

Appendices

Appendix A: Questionnaire

Demographic Questions

1. **What language(s) do you speak at home?**
   [Short answer]

2. **Did you already have a ChatGPT account prior to this workshop?**
   ☐ Yes
   ☐ No

ChatGPT Account Questions [Only displayed if the response to Question 2 is "Yes"]

1. **When did you create your ChatGPT account?**
   ☐ In 2022
   ☐ In 2023
   ☐ In 2024
   ☐ In 2025

2. **What version of ChatGPT do you use?**
   ☐ Free
   ☐ Paid

3. **For what tasks do you use ChatGPT?**
   ☐ For learning French
   ☐ For learning another language
   ☐ For things other than language learning

ChatGPT Use Questions [Only displayed in function of the responses selected for Question 3]

1. **How often do you use ChatGPT to learn French?**
   ☐ Daily
   ☐ Several times a week
   ☐ Once a week
   ☐ Several times a month
   ☐ Once a month
   ☐ Less than once a month

2. **How often do you use ChatGPT to learn languages other than French?**
   - ☐ Daily
   - ☐ Several times a week
   - ☐ Once a week
   - ☐ Several times a month
   - ☐ Once a month
   - ☐ Less than once a month

3. **How often do you use ChatGPT for things other than language learning?**
   - ☐ Daily
   - ☐ Several times a week
   - ☐ Once a week
   - ☐ Several times a month
   - ☐ Once a month
   - ☐ Less than once a month