Vol.5 Iss.6 2024 Communication & Education Review
https://doi.org/10.37420/j.cer.2024.093

From Material Generation to Vocabulary Acquisition: A New Model of Second Language Vocabulary Acquisition Driven by Large Language Model

Jing Li,Kun Li *
School of Foreign Studies, Zhaoqing University, Zhaoqing, China
***Corresponding author, e-mail: likun_qhd@126.com**

Abstract:
Along with the advancement of globalisation, foreign language proficiency has become a key skill, of which vocabulary acquisition is the core of language learning. The article aims to explore the effectiveness of Large Language Models (LLMs) generative materials in second language vocabulary acquisition and examine their effectiveness among learners of different language levels. Based on the comprehensible input hypothesis and the attention hypothesis, the study conducted a two-week vocabulary acquisition experiment with Japanese majors at a university in Guangdong Province, using experimental research method and quantitative analysis. The experiment compared the differences in vocabulary acquisition effects between learning materials generated by the Large Language Model and traditional instructional materials, and evaluated the effectiveness of the new vocabulary acquisition model by statistically analysing the scores of 39 students on immediate and delayed tests. The results of the study showed that the learning materials generated by the Large Language Model significantly outperformed the traditional materials in immediate vocabulary acquisition, but there was no significant difference in the effectiveness of the two methods in the delayed test. In addition, learners with different language levels did not show significant differences in vocabulary acquisition when using the learning materials generated by the Large Language Models

Keywords:
Large Language Model; second language vocabulary acquisition; immediate memory; delayed memory; personalised learning materials

1. Introduction

With the deepening development of globalisation, foreign language proficiency has gradually become one of the core competencies of global citizens. Policy documents of various countries constantly emphasise the importance of foreign language teaching, such as the Common European Framework of Reference for Languages (Supplementary Edition) (2018, 2020) and the Chinese English Proficiency Scale (2018), which explicitly focus on improving students' communicative competence and comprehensive application of language. In this context, vocabulary, as the core of language learning and the key to improving communicative

© 2024 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

competence, has received extensive attention from the applied linguistics community. Over the years, various theoretical frameworks such as "comprehensible input hypothesis", "attention hypothesis" and "deep processing theory" in the field of second language acquisition research have provided solid theoretical guidance for foreign language vocabulary teaching. However, the traditional second language vocabulary teaching However, traditional second language vocabulary teaching methods often rely on memorising word lists or inferring from contexts, and this kind of mechanistic learning is often difficult to adapt to the needs of modern learners. In recent years, with the continuous development of information technology, computer-assisted foreign language teaching (CALL), mobile-assisted foreign language teaching (MALL) and other technological means have been introduced into foreign language teaching, which have effectively improved the efficiency of vocabulary teaching. However, these technologies still have limitations in personalised learning resource production and instant feedback. With the emergence of large language models (LLMs), second language vocabulary acquisition has ushered in new opportunities.

Large Language Models are natural language processing technologies based on deep learning that can generate, understand and process human language. Since the launch of ChatGPT by OpenAI in 2022, various types of Large Language Models, such as Claude and NewBing, have been introduced one after another and have rapidly entered the field of foreign language teaching. Compared with traditional technologies, Large Language models are not only able to generate personalised second language learning materials for learners, but also provide instant feedback and exercises dynamically adjusted according to the learners' level, enhancing the interactivity and flexibility of the learning process. Based on this, this paper takes the "comprehensible input hypothesis" and "attention hypothesis" as the theoretical basis, and takes "the effect of second language vocabulary acquisition driven by the Large Language model" as the research object. Through experimental research and quantitative analysis, we will investigate whether the learning materials generated by the Large Language Model can effectively promote learners' vocabulary acquisition, and compare the differences between immediate and delayed memory in Large Language Model-driven vocabulary acquisition. In addition, it will also be analysed whether there are significant differences in the vocabulary acquisition effects demonstrated by the Large Language Model among learners with different language levels. This study aims to provide new ideas and methods for foreign language vocabulary teaching while expanding the research horizons of technology-assisted foreign language learning.

2. Literature review

2.1 L2 vocabulary acquisition

There is a rich history of research on second language vocabulary acquisition that focuses on how to facilitate vocabulary acquisition through effective input, attention and deep processing. Among them, the comprehensible input hypothesis (Krashen, 1985) advocates that learners acquire language progressively through "comprehensible input" (i+1), which is slightly above the current learner's language level and stimulates the natural growth of language proficiency. The attentional hypothesis (Schmidt, 1990) further explores the role of attention in second language vocabulary acquisition, i.e., learners must pay attention in order to become aware of new linguistic forms in the input and acquire these forms through conscious comprehension and processing processes. In addition, the deep processing theory proposed by Craik and Lockhart (1972) suggests

© 2024 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

that information is better remembered when it is processed more deeply. In second language vocabulary acquisition, deeper memory of vocabulary can be facilitated by more complex contexts, semantic associations, or repeated applications.

Based on the above theory, in recent years, researchers have successively introduced text annotation, multimodal resources, information technology and other means into the practice of foreign language teaching, relying on a new vocabulary teaching mode to enhance the effect of vocabulary acquisition of foreign language learners. Among them, Li Sihang (2016) attempted to combine mobile technology with multimodal resources and apply it to English vocabulary acquisition. It was found that this model has significant advantages over computer-assisted vocabulary acquisition (CALL), especially in terms of long-term retention of vocabulary memory. Cao Jiaxue (2014) experimentally compared the effects of three different annotation modalities on incidental vocabulary acquisition, and found that the choice of two English annotations was the most effective in the immediate test, while there was no significant difference among the three modalities in the delayed test. Cui Min (2015) applied text, static pictures and dynamic videos to vocabulary teaching respectively, and found that the presentation of static pictures and dynamic videos combined with text was more conducive to the long-term memory of vocabulary meanings, while there was no significant difference among the three presentation modes in terms of the memory of word forms.

Throughout the existing research, limited by technical bottlenecks, most of the research focuses on the effect of a traditional single technology, and has not yet explored how to effectively combine the use of a variety of technologies, and in practice, there are problems such as the high labour cost of producing vocabulary learning materials, and the difficulty of the content of the materials and the difficulty of the credibility of the control. The emergence of Large Language Modelling provides a new way of thinking to solve the above problems.

2.2 The Application of Large Language Modelling in Foreign Language Teaching and Learning

Large Language Models (LLMs) are deep learning-based artificial intelligence models that understand and generate human language. As the core of technology integration, they can combine multimodal resources, such as text, images and videos, to generate customised, dynamic and diversified foreign language learning materials according to learners' needs and levels, improving the degree of personalisation of learning and creating a more collaborative learning ecosystem. Thanks to large-scale corpora and advanced training algorithms, major language models represented by ChatGpt have made significant breakthroughs in language comprehension and generation, and have been applied to a number of foreign language teaching fields, such as translation, dialogue, writing assistance, etc., which have widely promoted the improvement of foreign language learning effects.Xiao (2023) et al. in their study designed a reading comprehension practice generation based on a large language model system based on a large language model to provide Chinese secondary school students with high-quality, personalised reading materials. The results of the study showed that the materials generated by the system were suitable for students' use and superior to the existing human-written materials.Cherednichenko (2024) et al. in their study explored the application of large-scale language modelling in foreign language learning in a distance online learning environment, and the results of the study showed that the application of large-scale language modelling can significantly improve students' language

© 2024 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

skills in four aspects: listening, speaking, reading and writing, and is important for translation Seßler (2023) and others developed and applied a tool called PEER (Paper Evaluation and Empowerment Resource) based on ChatGpt, which motivated students to improve their writing skills by providing them with comprehensive feedback and specific suggestions for improvement in writing foreign language essays.

The above studies provide important reference value and ideas for the introduction of Large Language modelling in foreign language teaching. However, looking at the existing results, most of the research focuses on the analysis of the improvement effect of listening, speaking, reading, writing and translating skills, but lacks attention to vocabulary acquisition, grammar application, cultural transfer and other sub-divisions, especially in the impact of the introduction of the large language model on the effect of vocabulary acquisition on foreign language learners is even more rare. In view of this, this paper intends to introduce a large language model into vocabulary acquisition in foreign language teaching based on the theories of comprehensible input hypothesis, attentional hypothesis, and deep processing theory, and carry out relevant teaching practice to test the effects of the new vocabulary acquisition model on immediate and delayed memory of L2 vocabulary, so as to provide a new way of thinking for foreign language teaching.

3. Research design

3.1 Research questions

Based on the comprehensible input hypothesis and the attention hypothesis, and in accordance with the needs of teaching practice, this paper attempts to answer the following questions: (1) Are learning materials generated by the Large Language Model effective in promoting learners' vocabulary acquisition? (2) What are the differences between immediate and delayed memory for vocabulary in the learning materials generated by the Large Language Model? (3) Is there a significant difference in the vocabulary acquisition effect of learning materials generated by the large language model among learners with different language levels?

3.2 Subjects of study

A total of 39 second-year students majoring in Japanese at a university in Guangdong Province were selected as subjects for this experiment, including 14 male and 25 female students. This group had completed basic Japanese knowledge at the beginning of the experiment, had a Japanese Language Proficiency Test (JLPT) N3 level, and had a vocabulary of about 2,500 to 3,000 words.

3.3 Research tools

The research tool used in this experiment is vocabulary test, and the experimental data mainly consists of vocabulary test scores based on the learning materials of the Large Language Models (hereinafter referred to as "LLMs mode scores") and vocabulary test scores based on the traditional learning materials (hereinafter referred to as "traditional mode scores"), each of which consists of immediate test scores and delayed test scores. (hereinafter referred to as "LLMs mode scores") and vocabulary test scores based on traditional learning materials (hereinafter referred to as "traditional mode scores"), each of which consists of immediate and delayed test scores. The Japanese vocabulary words covered in the study materials were all new words that

© 2024 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

had not yet been learnt. The vocabulary test consisted of: 1. vocabulary spelling (50 points), i.e., writing the kana corresponding to the Japanese words; 2. word-meaning linking (50 points), i.e., correlating the target words with their Chinese meanings. The vocabulary tests were conducted immediately after the classroom experiment and the scores obtained were used as immediate test scores, while the other test was conducted one week after the classroom experiment and constituted a delayed test score.

3.4 Study design

Before the beginning of the experiment, the teacher of the class selected 25 key words of different lexical forms in the vocabulary list of a certain unit of the textbook, after which all the words were entered into the ChatGpt dialogue box as a list and prompts were entered:

You are a senior Japanese teacher of Japanese nationality who is now teaching a reading class to Japanese students at a university in China. Use the following vocabulary to develop an interesting story as reading material for your students to help them understand and remember the new vocabulary. Requirements: 1. The story should be about 1,000 characters long; 2. The language and grammar used in the story should be at the JLPT N2 level; 3. All of the vocabulary listed must be used; 4. After each of the target vocabulary words, label it with a Japanese kana and its Chinese meaning, e.g., 携わる *(* たずさわる *) to take part in (* たずさわる *,* 参加 *); 5. Provide a translation of the Chinese version at the end of the story. The vocabulary list is as follows: (omitted)*

After completing the above operations, the system automatically generated a Japanese story that basically met the above requirements. After that, the teacher input the instruction again and asked the system to generate 2 stories with different plots again. After all the vocabulary learning materials were generated, the teacher selected one of the stories according to the students' receptive ability and partially modified the story by deleting or replacing new vocabulary other than the target words in the short text to reduce the cognitive load.

The vocabulary acquisition experiment was conducted in an offline classroom setting. In session 1, the instructor first asked students to read the vocabulary learning material generated by the Large Language Model for a time limit of 15 minutes. After reading, a target vocabulary test was administered using the experimental tool described in 3.3 to obtain the new model instant test scores. As a control, in session 2, the teacher selected another unit vocabulary list and text from the same textbook as learning materials, and also selected 25 new vocabulary words as a tool to test students' vocabulary acquisition. The text was approximately the same level of difficulty as the material generated by the large-scale language model, and the new target vocabulary was similar in difficulty to the target vocabulary used in the material generated by the large-scale language model. The target vocabulary test was administered after students had completed a 15-minute reading of the material. This portion of the score was used as the subject's immediate test score in the traditional vocabulary learning model.

One week after the end of the experiment, the teacher tested the vocabulary of the same group of students using the exact same vocabulary test materials to examine the delayed effects of vocabulary acquisition under the 2 different vocabulary learning modes.

© 2024 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

3.5 Data collection and processing

The vocabulary tests carried out in this experiment were completed online, and after the students completed the tests, the grades were generated automatically by the online testing system, and the grades were objective and authentic. After all the data statistics were completed, the data were analysed using spss27.0.

4. Experimental findings and discussion

4.1 Analysis of the effects of vocabulary acquisition driven by a large language model

At the end of the experiment, the data on vocabulary acquisition in the two different modes were statistically analysed by paired samples t-test, and the results of the experiment showed that the mean score of vocabulary acquisition in the new mode in the immediate test was 82.05 (SD=21.08), and that of the traditional mode was 50.64 (SD=23.74), and that the mean difference of the acquisition scores in the two modes (MD) was 31.41, and the difference was statistically significant by t-test ($t(38)=7.855$, $p<0.05$). Meanwhile, the result of effect size calculation (Cohen's d) was 1.40, indicating that there was a large effect difference between the two groups. The results of the above data indicate that the large language model-driven vocabulary acquisition model is significantly better than the traditional model in terms of short-term vocabulary acquisition effects.

In the delayed test, the mean score on the vocabulary test was 57.44 (SD=23.08) in the new mode; 56.03 (SD=23.63) in the traditional mode. The mean difference (MD) between the two modes was 1.41 and the t-test result was 0.612, which was not significant. Meanwhile, the effect size was calculated to be 0.06, further indicating that the difference in effect between the two groups was small and almost negligible. The above data indicate that the effect of the large language model-driven vocabulary acquisition model is similar to that of the traditional vocabulary acquisition model in terms of the effect of long term memory of vocabulary.

Table 1 Comparison of vocabulary acquisition effects between the large language model-driven vocabulary acquisition model and the traditional vocabulary acquisition model

	New model achievements		Traditional model achievements				
	M	SD	M	SD	MD	t (38)	p
Instant Testing	82.05	21.08	50.64	23.74	31.41	7.855**	<0.001
delay time test	57.44	23.08	56.03	23.63	1.41	0.511	0.612

4.2 Analysis of the effect of the large language model-driven vocabulary acquisition model on vocabulary acquisition among students of different levels of proficiency

On the basis of the above data analysis, the author divided the subject group into three groups according to their Japanese composite scores: high, medium and low, and further analysed the vocabulary acquisition effect of each group through repeated measures ANOVA with a two-factor mixed design, to examine whether

© 2024 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

there is a significant difference in the vocabulary acquisition effect of the large-scale language model generating materials among learners with different language proficiency levels .

The results of the analyses showed a significant main effect of vocabulary acquisition mode on vocabulary acquisition effects in the immediate test, which significantly influenced vocabulary acquisition effects regardless of learners' language level ($F = 60.621$, $p < 0.001$). The effect size of vocabulary acquisition mode was large (partial $\eta^2 = 0.627$), and this factor explained 62.7% of the total variance, indicating its strong influence in vocabulary acquisition. Meanwhile, the interaction effect between language level and vocabulary acquisition mode was not significant ($F = 0.668$, $p = 0.519$) and had a small effect size (partial $\eta^2 = 0.019$), suggesting that the effect of vocabulary acquisition between the different language level groups did not differ significantly between the two acquisition modes. In other words, the effect of vocabulary acquisition modes on the short-term effects of vocabulary acquisition does not vary according to learners' language levels. In summary, this study shows that the vocabulary acquisition mode driven by the Large Language Model significantly facilitates vocabulary acquisition in terms of short-term memory as a whole, and this effect is consistent across language level groups.

Table 2 Two-factor repeated measures ANOVA statistics (immediate test)

	F-value	P-value	bias η^2	mean square error
vocabulary acquisition pattern	60.621	<0.001	0.627	
Vocabulary Acquisition Patterns* Language Level	0.668	0.519	0.019	
error term variance				317.361

In the delayed test, the main effect of vocabulary acquisition mode was not significant ($F=0.289$, $p=0.594$), and the effect size was extremely small (biased $\eta^2=0.008$), which indicated that there was no significant difference in effect between the large language model-driven vocabulary memory mode and the traditional vocabulary memory mode in the delayed test, suggesting that there is no obvious superiority of the two methods in terms of long-lasting vocabulary memory. In addition, the interaction effect between the vocabulary acquisition model and the language level group approached significance ($F=3.014$, $p=0.062$), with a moderately small effect size (partial $\eta^2=0.143$). Although it did not reach the traditional level of significance ($p<0.05$), this result suggests that there may be some differences in vocabulary acquisition effects between the different language level groups in the two modes. Another noteworthy point is the large variance of the error term (mean square error = 134.402), suggesting that individual differences may have had a greater impact on vocabulary acquisition. In summary, the present study failed to find a significant main effect of vocabulary acquisition mode in delayed vocabulary acquisition, implying that the large language model-driven vocabulary memory mode and the traditional vocabulary memory mode have similar effects on long-term vocabulary retention.

© 2024 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

Table 3 Two-factor repeated measures ANOVA statistics (delayed test)

	F-value	P-value	bias η^2	mean square error
vocabulary acquisition pattern	0.289	0.594	0.008	
Vocabulary Acquisition Patterns* Language Level	3.014	0.062	0.143	
error term variance				134.402

Further post hoc comparisons showed that the mean scores under the large language model-driven vocabulary acquisition model were 75.38 (SD=16.64) for the high-level group, 58.85 (SD=16.60) for the medium-level group, and 38.08 (SD=19.42) for the low-level group; the corresponding scores under the traditional approach were 70.77 (SD=20.90), 52.69 (SD=17.87) and 44.62 (SD=24.96). The high and intermediate level groups improved by 4.61 and 6.15 points respectively, while the low level group decreased by 6.54 points. Although the differences between the groups did not show significance, it partly indicates that the vocabulary acquisition model driven by the Large Language Model is more facilitating for the vocabulary acquisition of the high and intermediate level learners, but no significant positive effect was seen in the low level group.

Table 4 Comparison of vocabulary acquisition effects between the new vocabulary acquisition model and the traditional vocabulary acquisition model in each level group

		New model achievements		Traditional model achievements		MD	t (12)	p
		M	SD	M	SD			
high level	Instant Testing	91.92	12.99	66.54	21.25	25.38	4.387	<0.001
	delay time test	75.38	16.64	70.77	20.90	4.615	0.938	0.367
mid-level	Instant Testing	85.38	17.26	47.49	21.85	37.69	5.142	<0.001
	delay time test	58.85	16.60	52.69	17.87	6.15	1.28	0.225
low level	Instant Testing	68.85	25.18	37.69	19.75	31.15	4.077	0.002
	delay time test	38.08	19.42	44.62	24.96	-6.54	-1.581	0.140

5. Discussion

5.1 Large language model-driven short-term vocabulary acquisition outperforms traditional models

The previous experimental data show that the novel vocabulary acquisition mode based on generative materials of the Large Language Model significantly outperforms the traditional text-memory or word list-memory modes in terms of immediate vocabulary recall effects. This data further validates the findings that generative learning materials help improve short-term vocabulary memory by relying on their highly interactive

© 2024 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

and personalised features (Joanne Leong et al., 2024). The reason for this is consistent with the fact that the personalised materials generated by the Large Language Model are able to customise the learning content according to the learners' needs, which can better attract students' attention, enhance the learning fun, emphasise the active construction of knowledge and learners' interactive participation, which is in line with the generative learning theory, and promote vocabulary memory and acquisition. Meanwhile, in the generative materials, text annotations were attached to each target word, which provided learners with comprehensible input materials, facilitated them to reduce the cognitive load by combining multi-dimensional information, such as context and pronunciation, and enhanced their effective attention to the target words, which helped them to quickly master the vocabulary meanings and pronunciation in a short period of time. This result also further validates the significance of the positive influence of text annotation on vocabulary acquisition (Cao Jiaxue, 2014).

On the other hand, the research data showed that the new vocabulary acquisition model did not show a significant difference from the traditional vocabulary acquisition model in terms of the delayed effect of vocabulary acquisition, and the difference between the effects of the two models was negligible. Although some research results suggest that generative learning materials may be beneficial for long-term memory (Al-Seghayer, K. 2005; Yun, 2011.), no significant difference between the two modes in delayed vocabulary acquisition was found in this study. This result may be related to the natural decline phenomenon of longterm memory mentioned in the current study (Schuetze, 2015), i.e., the gradual convergence of learning effects over time, regardless of the generative nature of the learning material. In the Large Language Model-driven model of second language vocabulary acquisition, learners' short-term memory for vocabulary is generated after completing the initial learning of the vocabulary due to the stimulation of interest and attention, but the lack of re-exposure to the target vocabulary learning material in the subsequent week may result in a single exposure that is insufficient to build a solid vocabulary memory. Meanwhile, text annotations and pronunciation cues may rely mainly on students' attention and shallow cognitive processing, and although they help students to quickly grasp vocabulary meanings in a short period of time, due to the lack of further deeper processing (e.g., use, association, or practice, etc.), these auxiliary information may fail to be transformed into deeper cognitive and autonomy-using abilities for the students (Nemati, A. 2009). This result also provides the researcher with a new idea that in future vocabulary teaching sessions, vocabulary output exercises such as speech, writing, and oral expression can be added to the provision of input generative information as a way of strengthening deeper processing of vocabulary and enhancing the effect of vocabulary acquisition (Li, S., 2021).

5.2 Large Language Modelling-driven Vocabulary Acquisition Effects Do Not Differ Across Language Level Groups

As mentioned earlier, the results of the two-factor repeated measures ANOVA showed a non-significant interaction effect between language level groups and vocabulary acquisition mode in terms of immediate memory for vocabulary with a small effect size, which implies that the vocabulary acquisition mode driven by the Large Language Model was able to have a significant positive effect on all language level groups. The reason for this may be related to the flexibility and individualised nature of the generated materials. In this experiment, the Large Language Model was able to adapt the difficulty of the materials according to the teacher's

© 2024 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

instructions, thus adapting to learners at different levels, and this adaptability may be the main reason for the non-significant difference in the learning effects between the different level groups. In particular, the learning materials generated by the Large Language model were able to provide learners with rich contextual cues and diverse memory paths through textual annotations and other cues, thus facilitating shallow vocabulary recall (Varol, 2017).

In terms of delayed memory, the interaction effect between vocabulary acquisition mode and language level group was close to significant, though, suggesting that there may be differences in the delayed effects of vocabulary acquisition across language level groups. The large variance of the error term likewise suggests that individual differences may play a large influential role in the process of vocabulary memory. Further analyses showed that the difference in performance between the two modes of vocabulary acquisition in the delayed test was 4.615 for the high-level group, 6.15 for the intermediate-level group, and -6.54 for the low-level group.These details suggest that, although the differences did not reach the level of significance, the material generated by the large language model was slightly more helpful to the high- and intermediate-level groups in the delayed test, while it was not as effective as the traditional method for the low-level group. Instead, it was less effective than the traditional method. This may reflect the fact that low level learners have less vocabulary and language knowledge reserves, lower language proficiency, rely more on the annotations and pronunciation information of words, lack autonomous use of vocabulary and deeper memory, and fail to connect the target words with the existing knowledge network (Ma, 2008). At the same time, it may have encountered a higher cognitive load when processing complex texts, resulting in a higher burden in delayed memory and failing to fully assimilate the complex materials generated by the Large Language Models Future research should focus on exploring the adaptive improvement of materials generated by large language models for low-level learners, and may consider adjusting the difficulty of the materials or providing more guidance to help such learners to better enhance their vocabulary long-term memory effects.

6. Conclusion

In this study, based on the comprehensible input hypothesis and the attention hypothesis, we designed and developed vocabulary learning materials based on the Large Language Model, and explored their immediate and delayed effects in second language vocabulary acquisition for learners of different language levels. The results of paired-samples t-test and two-factor repeated measures ANOVA indicated that the vocabulary learning materials generated by the Large Language Model showed significant advantages in immediate vocabulary acquisition. Regardless of students' language level, the immediate test scores in the novel model were significantly higher than those in the traditional glossary memorisation model. The study concluded that the interactivity, fun and accompanying text annotations of the personalised learning materials generated using the Large Language Model helped to attract learners' attention and promote short-term vocabulary acquisition. However, the difference between the two modes was not significant in the delayed test and did not show significant differences across language level groups, which resulted in the suggestion that a single session of generative learning materials was not significant in terms of long-term vocabulary memory retention, and that more vocabulary output tasks with repetitive use may be needed to maintain the long-term effect. It is worth mentioning that under the novel model, the performance of low-level learners in delayed memory is poorer,

© 2024 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

even lower than that of the traditional method, which also suggests that we need to take into account the cognitive load of low-level learners when applying complex materials generated by the large language model.

This study expands the boundaries of the application of the Large Language Model in L2 vocabulary acquisition in terms of theory, verifies the immediate effects of the Large Language Model's generative materials, supports the positive effects of generative materials and personalised learning content on vocabulary acquisition in a short period of time, strengthens the association between generative learning and the attentional hypothesis, and expands the previous framework of vocabulary acquisition that relies mainly on static textual input. Meanwhile, the study systematically explores for the first time the applicability of the Large Language Model to learners at different language levels, providing theoretical support for the future application of the Large Language Model to learners at different levels. In terms of practice, this study verifies the effectiveness of the learning materials generated by the Large Language Model, innovatively proposes new ideas to solve the problems of high cost of personalised resource production and insufficient feedback in the previous technology-assisted vocabulary teaching, and provides a new teaching tool for foreign language teachers and educational institutions.

Although this study experimentally verified the immediate effect of the Large Language Model in L2 vocabulary acquisition, it still suffers from research deficiencies such as limited sample size, short research period, a single vocabulary learning model, and learners' individual differences not fully considered due to the research conditions. Future research can further expand the sample size to cover more languages and learners from different backgrounds to further verify the broad applicability of the large language model to vocabulary acquisition. Meanwhile, future studies can also further optimise the application of the model by trying to integrate various teaching techniques and tools, and use the Large Language Model in conjunction with multimodal resources and output tasks, so as to enhance learners' deeper processing of vocabulary and to promote the long-term memory of second language vocabulary. In addition, although this study analysed subgroups by language level, it still failed to explore in depth other individual factors that may affect vocabulary acquisition, such as motivation, cognitive style and learning strategies. Future research could also try to further explore how individual differences in learners affect the use of the Large Language Model, and design more personalised learning strategies for different types of learners.

Funding

This paper was supported by the following three projects:

[1] Zhaoqing Institute of Educational Development Project (No. ZQJYY2023213)

[2] Guangdong Provincial Education Science Planning Project (Higher Education) (No. 2022GXJK357)

[3] Zhaoqing University Higher Education Teaching Reform Project (No.zlgc2024042)

References

[1]Al-Seghayer, K. (2005). The effect of multimedia annotation modes on L2 vocabulary acquisition. In Re-

© 2024 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

search in technology and second language education: Developments and directions . Information Age Publishing. 133-151
[2]Cao, J., & Song, J. (2014). The effects of different annotation methods on incidental vocabulary acquisition. Journal of Foreign Languages, 1(1), 117-120.
[3]Cherednichenko, O., Yanholenko, O., Badan, A., Onishchenko, N., & Akopiants, N. (2024). Large language models for foreign language acquisition. Computational Linguistics Workshop at 8th International Conference on Computational Linguistics and Intelligent Systems (pp. 520-535). Computational Linguistics Workshop.
[4]Council of Europe. (2018). Common European framework of reference for languages: learning, teaching, assessment (Companion volume with new descriptors). Council of Europe.
[5]Council of Europe. (2020). Common European framework of reference for languages: Learning, teaching, assessment (Companion volume). Common European framework of reference for languages: Learning, teaching, assessment (Companion volume).
[6]Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: a framework for memory research. Journal of Verbal Learning and Verbal Behavior, 11(6), 671-684.
[7]Cui, M., & Zhou, C. (2015). A study on the effectiveness of multimedia presentation in the direct learning of foreign language vocabulary. journal of PLA College of Foreign Languages, (03), 88-95.
[8]Krashen, S. (1985). The input hypothesis: Issues and implications. Longman.
[9]Leong, J., Pataranutaporn, P., Danry, V., Perteneder, F., Mao, Y., & Maes, P. (2024). Putting things into context: Generative AI-enabled context personalisation for vocabulary learning improves learning motivation. In Proceedings of the CHI Conference on Human Factors in Computing Systems (Article 677, pp. 1-15). Association for Computing Machinery.
[10]Li, S.-H., & Gao, Y. (2016). An empirical study on the effectiveness of mobile technology-assisted foreign language teaching on English vocabulary acquisition. foreign Language Community, (04), 73-81.
[11]Li, X., & Lei, L. (2021). The effect of output tasks on incidental vocabulary acquisition in English reading: a meta-analysis based on domestic related studies. Journal of PLA College of Foreign Languages, (06), 18-26.
[12]Ma, Q. (2008). Empirical CALL evaluation: the relationship between learning process and learning outcome. CALICO Journal, 26(1), 108-122.
[13]Ministry of Education Examination Centre. (2018). Chinese English proficiency scale. Higher Education Press.
[14]Nemati, A. (2009). Memory vocabulary learning strategies and long-term retention. International Journal of Vocational and Technical Education, 1(2), 14- 24.
[15]Schmidt, R. (1990). The role of consciousness in second language learning. Applied Linguistics, 11(2), 129-158.
[16]Schuetze , U. (2015). Spacing techniques in second language vocabulary acquisition: short-term gains vs. long-term memory. Language Teaching Research, 19(1), 28-42.
[17]Seßler, K., Xiang, T., Bogenrieder, L., & Kasneci, E. (2023). PEER: Empowering writing with large language models. In O. Viberg, I. Jivet, P. Muñoz-Merino, M. Perifanou, & T. Papathoma (Eds.), Responsive and sustainable educational futures. EC-TEL 2023. Lecture Notes in Computer Science (Vol. 14200). Springer.
[18]Varol, B. (2017). Annotation and working memory in second language reading, incidental vocabulary

© 2024 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)

learning, and perceived cognitive load. Archives and Documentation Center Digital Archives,12(1),15-20
[19]Xiao, C., Xu, S. X., Zhang, K., Wang, Y., & Xia, L. (2023). Evaluating reading comprehension exercises generated by LLMs: A showcase of ChatGPT in education applications. In Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (pp. 610-625). Association for Computational Linguistics.
[20]Yun, J. (2011). The effects of hypertext glosses on L2 vocabulary acquisition: a meta-analysis. Computer Assisted Language Learning, 24(1), 39-58.

© 2024 by the author(s); licensee Mason Publish Group, this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (http://creativecommons.org/licenses/by/4.0/)