INNOVATION IN LANGUAGE LEARNING International Conference

A Large Language Model-Infused Platform for Blending Spaced Repetition and Immersion in Mandarin Vocabulary Acquisition

Jacob Hume

University of Michigan, United States of America

Abstract

This work outlines the design and implementation of a large language model (LLM)-infused, ARID (availability, retention, immersion, data)-based digital flashcard system for traditional and simplified Mandarin vocabulary acquisition. Learners efficiently input multimedia — text, audio, spreadsheet, image — into a uniform translation interface (available on both mobile and desktop), which proceeds to automatically generate in the background flashcards for the popular open-source software Anki. These flashcards are complete with images, audio, example usage, mnemonics for meaning (via etymology) and writing (via radical decomposition), classifiers, related words, "how-to-write" gifs, etc. Retention is achieved via the spaced repetition of flashcards using Anki's scheduling algorithm. This may be accomplished per convention via Anki's desktop, mobile, or web app. It may also be accomplished via a custom web client which creates and reviews Anki cards en masse *during immersive practice with an LLM conversation partner. This web client consists of four modes, each of which will automatically create flashcards from any unknown content the learner encounters during conversation. "Translate" handles translation prompts too nuanced for 'classical' translation services. "Learn" mini-tutors the user on new cards. "Review" drills recently learned words via dynamically generated exercises. "Converse" allows the learner to engage in freeform conversation on a topic of their choice. The LLM will implicitly incorporate long-term review words into each interaction, allowing for the batch review of old Anki flashcards in the context of novel conversation. A snowball effect results — additional flashcards often get made when review happens, but the batch review process ensures that review sessions can nevertheless be completed efficiently. Lastly, custom study routines (available both in Anki and the custom web client) can be dynamically created by the user based on semantic, syntactic, orthographic, or exam data. We believe the system presented establishes an intimate interface between spaced repetition and immersion in language learning, allowing users to integrate state-of-the-art spaced repetition scheduling algorithms into their learning routine without sacrificing time spent practicing in an immersive environment.*

Keywords: *Computer-Assisted Language Learning (CALL), Spaced Repetition, Large Language Models (LLMs), Mandarin Vocabulary Acquisition, Immersive Learning*

1. Introduction

The recent popularization of general-purpose large language models (LLMs) has led to surging interest in using the technology for computer-assisted language learning (CALL) tasks [1, 2, 3], provided the target languages are well-represented in training data [4]. Regimes of CALL literature in which LLMs have seen adoption include automated language assessment [5, 6], simulated speaking practice [7, 8], and mixed reality [9]. Analogously, this work focuses on using LLMs for second language vocabulary acquisition tasks, focusing on Mandarin vocabulary as a representative example.
A number of technology-mediated approaches to second language vocabulary acquisition have found empirical success, including approaches based on contextual quizzing [10], multimedia glossing [11], and — perhaps most popular — the spaced review ('spaced repetition') of atomic concepts [12, 13, 14]. A given atomic concept is often suggestively dubbed a 'digital flashcard' or just 'card' in this context, a convention we adopt here even when no explicit flashcard representation is involved. In this work, we present the design and implementation of an LLM-infused software for Mandarin vocabulary acquisition that not only incorporates the aforementioned methodologies but is often compatible with existing tools implementing those methodologies. Its design centers on what will be hereon termed the ARID (**a**vailability, **r**etention, **i**mmersion, **d**ata) principles.

Availability. Regardless of location and medium of interaction, learners should be able to access the card creation process without sacrificing any present immersion. This is accomplished via Alfred

(MacOS) and Siri (iOS) translation workflows which automatically parse text, screenshot, audio, and photo input into flash cards. This is the content of Section 2.1.

Retention. Considering literature on spacing effects in knowledge formulation [15, 16, 17], a spaced repetition system (SRS) should be integrated to ensure that items are reviewed with optimal timing. The most popular open-source spaced repetition software, Anki [18], acts as a backend for managing card data and scheduling reviews based on the state-of-the-art FSRS algorithm [19]. Section 2.2 details this process.

Immersion. A learner should be able to practice in an immersive environment — if desired, they should be able to learn and review in conversational settings. To this end, an LLM interface is provided which provides in-depth translations, micro-lessons for new words, dynamic translation and cloze deletion drills for review words, or otherwise freeform conversation concerning on a topic chosen by the learner. The latter mediums incorporate 'batch reviews' of aging cards naturally incorporated into conversation, and 'keep track' of new words and questions the learner asks throughout the conversation to be automatically made into flashcards after. Details are discussed in Section 2.3.

Data. A learner should have reach toward extensive resources regarding the learning contents, allowing them to eventually settle on a practice routine that is personal and flexible. Numerous special tags and fields exist from which cards may be organized. This is the discussed in Section 2.4.

2. Method

We outline the proposed method, using the ARID principles as an expository template. The premise is as follows: for a given user, the system operates on a database of *words*, treated as the atomic concepts ('cards' in spaced repetition parlance) to be retained through continuous engagement with the software. The goal is to have the user *learn* a consistent number of new words for the first time on each day and *review* a consistent number of previously learned words on each day, in a manner roughly synced with the forgetting curve(s) studied in the spaced review literature [12, 13, 14]. Provided with a set of words to learn and review, our method employs LLMs and/or multimedia flashcards to immersively and efficiently facilitate the learning and reviewing tasks. To determine *which* cards to learn/review, we employ the FSRS spaced repetition algorithm introduced in [19]. To *have* cards to learn/review in the first place, we provide a diverse array of methods for creating rich multimedia cards from sparse, flexible input, a task also aided by LLMs. All cards are stored and tracked using the popular open-source digital flashcard platform Anki [18]. We use Streamlit [20] to host the web component for interaction with LLMs.

Remark. The term 'word' is used here for the sake of concreteness. However, the actual prompt provided to the LLM when segmenting input text is to partition according to 'atomic meaningful concepts' — that is, into meaningful portions of the text that would lose all meaning if divided up further. This is equivalent to seeking actual words in the vast majority of cases, but not all — idioms, for example, often qualify to be made into cards as well.

2.1 Availability: Efficiently Creating Cards

The main goal for card creation is to require as little information from the user as possible. To this end,

- The card creation process can be accessed on both mobile and desktop platforms. We implement on iOS and MacOS here, but analogous implementations on other operating systems would be straightforward to perform.
- All that is required to create card(s) is to provide an indefinite amount of text, audio, or image via one of many accepted input methods.
- Any time a user looks up a new word or encounters a new word 'in the wild', our system provides a way to automatically create card(s) from the experience.

In summary, the process of creating a card, in principle, takes no longer than the process of looking up or otherwise encountering a new word.

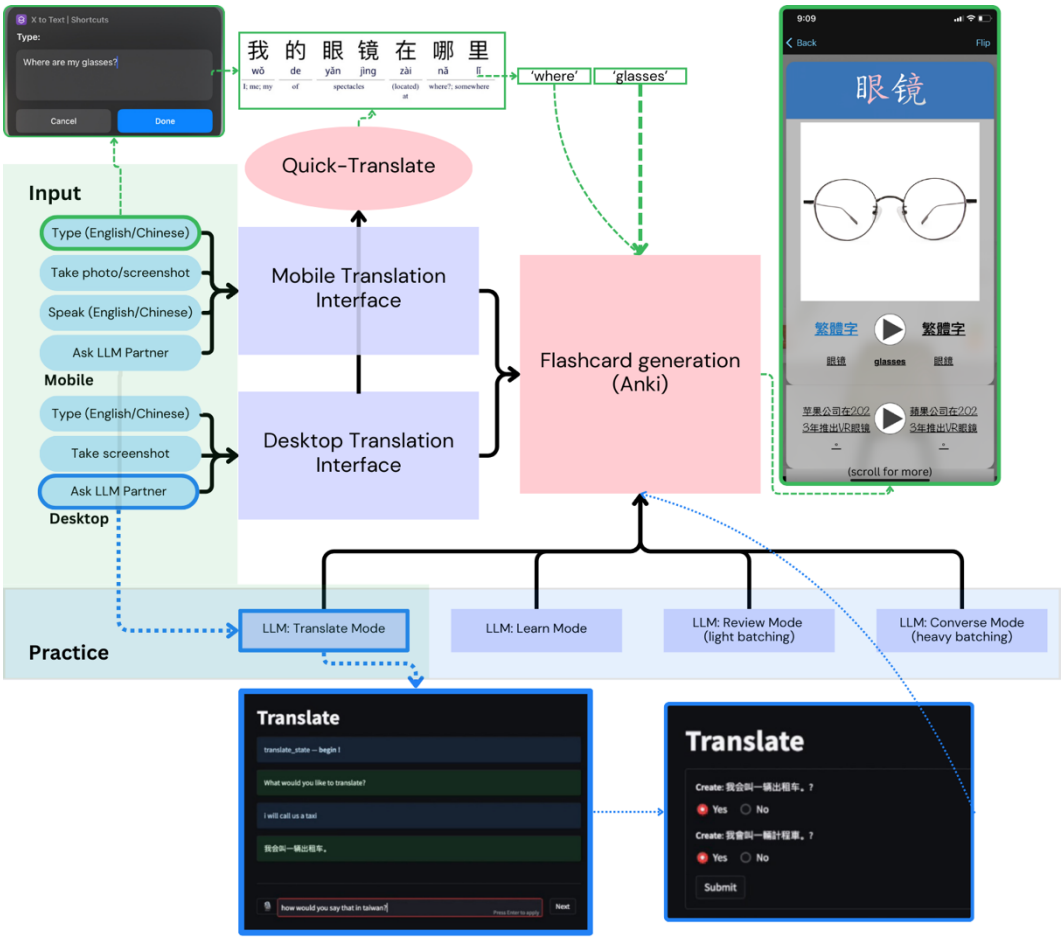

Fig. 1. A global schematic of card creation via the proposed method. Two representative information flows are depicted. The first information flow (green) illustrates a possible trajectory of new words on iOS. A text query is provided by the user, for which a rough translation using the built-in service in iOS shortcuts is given. The original query is parsed into words, and for every word without a flashcard a flashcard is made (here, the final 'glasses') card is shown. The second information flow (lavender) illustrates a possible trajectory of new words on desktop. In this case, an LLM helps the user obtain their desired translation. When the user clicks the *Next* button, the LLM provides potential excerpts to create flashcards from; upon approval, the excerpts are segmented as before and for all words without a card a card is made.

Creating Cards on iOS. An iOS shortcut provides translations and pinyin for Chinese or English input. These translations are just meant to be heuristics; they do not actually make it into the end card. Instead, the input is uploaded to an online database (here, we just use a Google Sheets) along with some metadata, such as the source from which data was uploaded. The next time Anki has internet access, the Anki add-on will create (in the background) a digital flashcard for each word in the input for which it has not made a flashcard before. If the input is sufficiently long, it will use that sentence as the flashcard's example sentence (the content of digital flashcards will be detailed in Section 2.2).

Accepted input methods are

- Typed text
- Clipboard
- Most recent screenshot
- Take screenshot
- Camera

- Camera roll
- Speak English
- Speak Chinese
- Talk with an LLM

Creating cards on MacOS. An Alfred Workflow [21] provides real-time translation heuristics for typed input. Upon pressing of the return key, the input is sent to the Google Sheet. From there it works the same as iOS. It is also capable of parsing screenshots for text and creating cards accordingly.

Creating cards from LLM conversations. In any conversation with the LLM interface (including the learning and reviewing panels in addition to translation, as we will discuss in Section 2.3), pressing the *Next* button will prompt the LLM to review the conversation up to the previous time *Next* was pressed and find text which it thinks 'should be flashcarded'. A common example of text that 'gets flashcarded' in this manner would be when a learner doesn't recognize a word in one of the LLM's sentences and asks for clarification. A more detailed discussion will be held in Section 2.3.

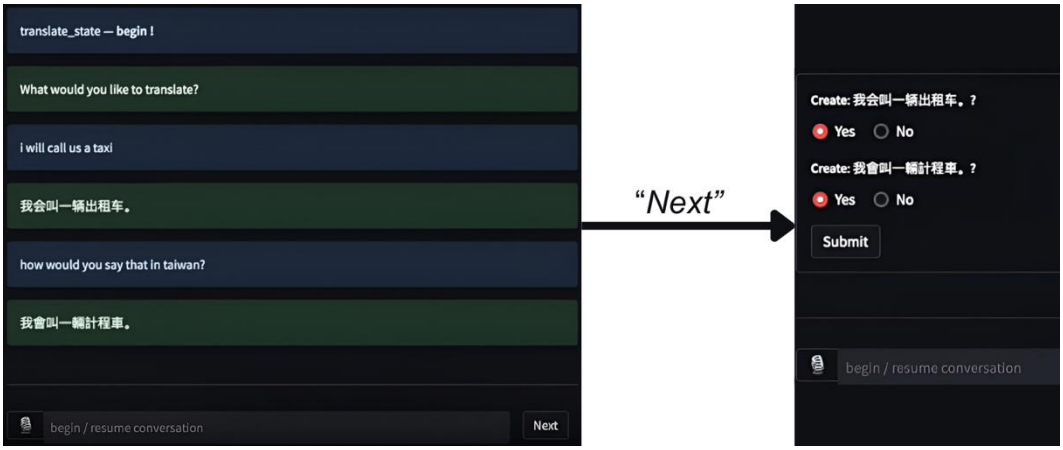

Fig. 2. A close-up example of creating new cards based on conversation with an LLM in "translation mode" (see Figure 1 for a contextualized version). New cards can also be made during conversations in "learn mode" and "review mode"; see Section 2.3.

Creating cards *en masse* using Google Sheets It is also possible to manually fill in Google Sheet entries. This is especially useful e.g. when inputting a vocabulary list for a new textbook chapter.

	A	B
1	Text	Source
2	這是老王,他開了一家公司,公司有很多員工,也賺了很多錢,員工也很喜歡他。	CC3-Lesson-5
3	他成功(地)開了一家公司。	CC3-Lesson-5
4	他管理公司管理得很成功。	CC3-Lesson-5
5	這裡的網路不穩定,上線一直不成功。	CC3-Lesson-5
6	下個月要去台灣留學了,我很興奮。	CC3-Lesson-5

Figure 3. An example Google Sheet excerpt. The rows are asynchronously and automatically segmented into words, and a card is made for all words without one. Each row depicted corresponds to a sentence taken from the reading in Lesson 5 of [21].

2.2 Retention: Structuring the Content of Cards

The Google Sheet is algorithmically parsed row-by-row, where a given row consists of a 'text' entry and 'source' metadata. The text comprising a row is then parsed into words, and a new card is created for each word that does not yet have one. The fields filled in by the algorithm for each card are

illustrated in Figure 4, visually compressed in light of space limitations. We note that the vocabulary word (situated the top of the card) is animated to illustrate stroke order, something that is not captured by the static image.

For each card, several tags are automatically added as well, which allow for the customization of review sessions. Common examples include choosing to study all cards with a specific source (for example, all vocabulary from a specific chapter of a textbook), studying words corresponding to a specific topic, or studying words according to a specific radical; this will be the focus of Section 2.4.

Our system uses Anki's FSRS scheduler [19] to optimally determine when cards should be learned and reviewed. The order in which cards are presented for learning and review is made accordingly.

Fig. 4. The backside of a flashcard created for the word 'thunder' by the proposed method. (The frontside of a flashcard is either the word in Chinese, the word in English, audio for the word, or the image.) English, simplified (left column) Chinese, traditional (right column) Chinese, and audio are provided for the word, an example sentence, synonyms, antonyms, classifiers, related words, and usages. English, pinyin, and a mnemonic may be accessed by clicking underlined text. If the original input used to create the card is a full sentence, then that sentence will be used as the example sentence. The word at the top of the flashcard is animated to demonstrate stroke order. Other fields which are filled in by our method but not shown here are phonetic regularities, radicals, and a decomposition of the word in terms of these.

2.3 Immersion: Conversationally Reviewing Cards with an LLM

One can, of course, learn and review cards from within the Anki software itself. However, one can also learn and review immersively and in batch using the LLM-integrated web app that we turn to discuss now. Keeping in line with our stipulation in Section 2.1 that the transition from encountering a new word to creating a card for said new word should be seamless, it also is possible semi-automatically create new cards based on LLM interactions.

Learning new cards with an LLM.

The conversation takes the form of a mini-tutoring lesson structured as follows:

- The LLM introduces the new word, its etymology, and perhaps some tricks for remembering it.
- The user asks questions and practices usage.
- Throughout, the LLM will incorporate long-term review cards into the dialogue as 'batch review'.

Once the user feels as though they have understood and learned the new word, they click the *Next* button. They are prompted to rate the difficulty of the new card as well as the infused long-term review cards according to Anki's rating system, and to multiselect from a provided list any incidentally novel words involved in the conversation for which that they would like the system to create new cards. The ratings are then used by Anki to curate the content of future sessions.

Reviewing cards with an LLM. Here, the learner drills recently learned words and batch-reviews long-term words. Drills may include dynamic translation exercises or cloze deletions, though we remark that in principle any LLM prompt may be used. Laced throughout these exercises are uses of long-term words. The *Next* button functions identically to before.

Conversing with an LLM. It is also possible to engage in freeform conversation with an LLM without any explicit review target, but nevertheless with the functionality to create new cards out of unknown words encountered during the conversation. If conversation is tailored toward a specific topic, the program will naturally incorporate review cards on that topic into the conversation.

Translating with an LLM. One can also create cards directly using an LLM, by conversing with it regarding how to translate content; this was alluded to in Section 2.1. The benefit over the standard approach to translation lies in the opportunity to ask clarifying followups, e.g. regarding grammatical usage, etymology, regional differences, etc. Pressing *Next*, as usual, will allow the program to create new cards out of the conversation.

Fig. 5. An example of reviewing a card with the proposed method via conversation with an LLM. In this example, the LLM has been instructed to dynamically provide cloze deletion prompts. The user is able to ask about the meaning of a word used by the LLM before attempting to answer the prompt. Pressing *Next* allows the user to rate the difficulty of the review (the same as they would in Anki), as well as approve the creation of a new card from the word they failed to recognize.

2.4 Data

Using our system, a learner has reach toward extensive resources regarding the learning contents, allowing them to eventually settle on a practice routine that is personal and flexible. For example, custom study schedules may be set within Anki using the multitude of tags automatically generated by the software (cf. Table 1). These can be used to arrange custom study sessions (e.g., 'study all verbs in HSK 7-9 with the radical 言'), steer LLM conversations ('let's review cards through a conversation about `art_and_literature`), etc. See Table 1.

Table 1. Types of tags automatically attached to Anki cards using the proposed method.

Category	Examples
`Source`	'Integrated Chinese Textbook 4, Chapter 7', 'added from iPhone'
`Level`	'HSKv2 level 4', 'HSKv3 level 6', 'not in HSKv3'
`Frequency`	'in top 40% of most frequently used words'
`Semantic`	'geography and landmarks', 'Chinese mythology and folklore', 'etiquette and social norms'
`Classifier`	'个', '台', '辆', '座'
`Orthographic`	'simplified contains radical 乚', 'traditional contains radical 車'
`Syntactic`	'adjective', 'is a grammar point', 'idiom', 'is a measure word'
`Custom`	(User-defined)

3. Discussion

We presented a novel approach to synthesizing immersion and spaced repetition in a CALL system for second-language Mandarin vocabulary acquisition, using large language models and the FSRS scheduling algorithm. The approach automates all scaffolding of the learning process by creating rich digital flashcards from sparse multimodal input, creating cards automatically when a user looks up a translation or encounters a new word in during conversation with an LLM, and immersively reviewing cards in batch and with theoretically principled timing.

REFERENCES

[1] Liu, W., Y. Lin, and Z. Yu, "A bibliometric analysis of artificial intelligence chatbots in language education," *Social Education Research*, pp. 59–78, 2024.

[2] Kovačić, A., G. Bubaš, *et al.*, "Using conversational artificial intelligence for online learning activities in english for specific purposes: A pilot study of students' experiences with Bing chat," in *Conference proceedings. Innovation in language learning 2023*, 2023.

[3] Hong, W. C. H., "The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research," *Journal of Educational Technology and Innovation*, vol. 5, no. 1, 2023.

[4] Lankford, S., H. Afli, and A. Way, "Transformers for low-resource languages: Is Féidir linn!" *arXiv preprint arXiv:2403.01985*, 2024.

[5] Caines, A. *et al.*, "On the application of large language models for language teaching and assessment technology," *arXiv preprint arXiv:2307.08393*, 2023.

[6] Jeon, J., "Chatbot-assisted dynamic assessment (CA-DA) for L2 vocabulary learning and diagnosis," *Computer Assisted Language Learning*, vol. 36, no. 7, pp. 1338–1364, 2023.

[7] Talkpal.AI, "Home." https://talkpal.ai, 2023.

[8] Hidayatullah, E., "The impact of talkpal. AI on english speaking proficiency: An academic inquiry," *Journal of Insan Mulia Education*, vol. 2, no. 1, pp. 19–25, 2024.

[9] Chafiq, N., I. Elimadi, and M. Talbi, "Metaverse and artificial intelligence: Towards a pedagogical revolution in hybrid language teaching in a university context (short paper)." in *AIxEDU@ AI* IA*, 2023.

[10] Groot, P. J., "Computer assisted second language vocabulary acquisition," 2000.

[11] Gorjian, B., S. R. Moosavinia, K. Ebrahimi Kavari, P. Asgari, and A. Hydarei, "The impact of asynchronous computer-assisted language learning approaches on english as a foreign language high and low achievers' vocabulary retention and recall," *Computer Assisted Language Learning*, vol. 24, no. 5, pp. 383–391, 2011.

[12] Chukharev-Hudilainen, E. and T. A. Klepikova, “The effectiveness of computer-based spaced repetition in foreign language vocabulary instruction: A double-blind study,” *calico journal*, vol. 33, no. 3, pp. 334–354, 2016.

[13] Nakata, T., “English vocabulary learning with word lists, word cards and computers: Implications from cognitive psychology research for optimal spaced learning,” *ReCALL*, vol. 20, no. 1, pp. 3–20, 2008.

[14] Zhu, Y., A. S. Fung, and H. Wang, “Memorization effects of pronunciation and stroke order animation in digital flashcards,” *CALICO Journal*, vol. 29, no. 3, pp. 563–577, 2012.

[15] Yuan, X., “Evidence of the spacing effect and influences on perceptions of learning and science curricula,” *Cureus*, vol. 14, no. 1, 2022.

[16] Woźniak, P. and E. Gorzelańczyk, “Optimization of repetition spacing in the practice of learning,” *Acta neurobiologiae experimentalis*, vol. 54, no. 1, pp. 59–62, 1994.

[17] Murre, J. M. and J. Dros, “Replication and analysis of ebbinghaus’ forgetting curve,” *PloS one*, vol. 10, no. 7, p. e0120644, 2015.

[18] Anki, “Anki - powerful, intelligent flashcards.” https://apps.ankiweb.net, 2023.

[19] Ye, J., J. Su, and Y. Cao, “A stochastic shortest path algorithm for optimizing spaced repetition scheduling,” in *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, 2022, pp. 4381–4390.

[20] Adrien Treuille and Streamlit Team. Streamlit: An open-source app framework for machine learning and data science teams, 2019. Version 1.0.

[21] Andrew Pepperrell and Vero Pepperrell. Alfred: Productivity app for macos, 2010. Version 5.0

[22] National Taiwan Normal University. A Course in Contemporary Chinese, Volume 3. Linking Publishing Company, Taipei, Taiwan, 2016.