VocabEncounter: NMT-powered Vocabulary Learning by Presenting Computer-Generated Usages of Foreign Words into Users' Daily Lives

Riku Arakawa[*]
Carnegie Mellon University
Pittsburgh, USA
rarakawa@andrew.cmu.edu

Hiromu Yakura[*]
University of Tsukuba / National Institute of Advanced Industrial Science and Technology (AIST)
Tsukuba, Japan
hiromu.yakura@aist.go.jp

Sosuke Kobayashi
Tohoku University
Sendai, Japan
Preferred Networks
Tokyo, Japan
in2400@gmail.com

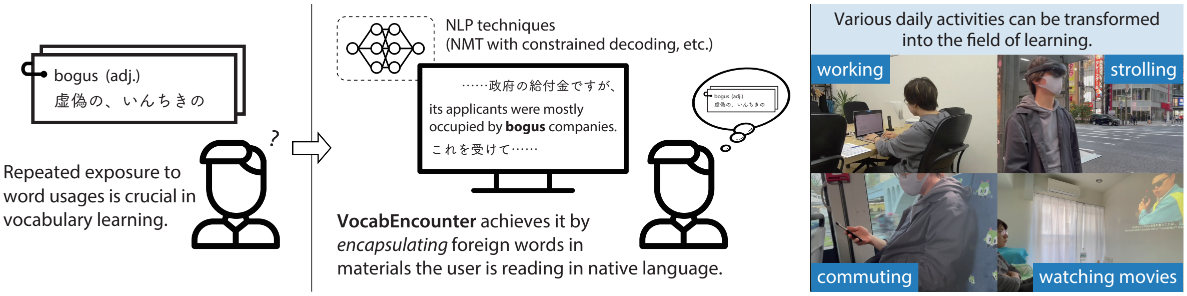

bogus (adj.)
虚偽の、いんちきの
?
Repeated exposure to word usages is crucial in vocabulary learning.
NLP techniques (NMT with constrained decoding, etc.)
……政府の給付金ですが、
its applicants were mostly occupied by **bogus** companies.
これを受けて……
bogus (adj.)
虚偽の、いんちきの
VocabEncounter achieves it by *encapsulating* foreign words in materials the user is reading in native language.
Various daily activities can be transformed into the field of learning.
working
strolling
commuting
watching movies

Figure 1: VocabEncounter enables users to encounter foreign words by presenting their usages generated using NLP techniques into materials the user is reading. Using this approach, the user can transform their daily life into the field of vocabulary learning.

ABSTRACT

We demonstrate that recent natural language processing (NLP) techniques introduce a new paradigm of vocabulary learning that benefits from both micro and usage-based learning by generating and presenting the usages of foreign words based on the learner's context. Then, without allocating dedicated time for studying, the user can become familiarized with how the words are used by seeing the example usages during daily activities, such as Web browsing. To achieve this, we introduce *VocabEncounter*, a vocabulary-learning system that suitably *encapsulates* the given words into materials the user is reading in near real time by leveraging recent NLP techniques. After confirming the system's human-comparable quality of generating translated phrases by involving crowdworkers, we conducted a series of user studies, which demonstrated its effectiveness on learning vocabulary and its favorable experiences. Our work shows how NLP-based generation techniques can transform our daily activities into a field for vocabulary learning.

CCS CONCEPTS

• **Human-centered computing** → **Natural language interfaces**; • **Computing methodologies** → **Machine translation**; • **Applied computing** → *Interactive learning environments*.

KEYWORDS

natural language processing, neural mechanical translation, vocabulary learning

ACM Reference Format:
Riku Arakawa, Hiromu Yakura, and Sosuke Kobayashi. 2022. VocabEncounter: NMT-powered Vocabulary Learning by Presenting Computer-Generated Usages of Foreign Words into Users' Daily Lives. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3491102.3501839

1 INTRODUCTION

Acquiring vocabulary is of paramount importance in learning a foreign language as it is fundamental knowledge to use and understand expressions of the language [1, 4]. Given that, a number of educational systems have been introduced for vocabulary learning, reflecting the expansion of Web and mobile technology, such as online vocabulary games [88] and mobile apps for flashcards [32, 59]. Also, HCI researchers have leveraged various interaction techniques for vocabulary learning, from app notifications [23] to multimedia retrieval [90] or mixed reality [85].

[*] These authors contributed equally and are ordered alphabetically.

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9157-3/22/04.
https://doi.org/10.1145/3491102.3501839

There are two main directions for supporting vocabulary learning: *micro* learning [30, 40] and *usage-based* learning [49, 90]. The strategy of micro learning was introduced to solve the dilemma between the necessity of repeated word exposure to enhance memory and the inconvenience of allocating dedicated time for vocabulary learning, as Dingler et al. [23] implemented a mobile app enabling instant learning. Some recent techniques present the words in a manner that is grounded in a user's context to learn them in small time segments effectively; for example, Edge et al. [26] developed a smartphone app that presents a short vocabulary quiz of relevance to the user's location. On the other hand, whereas the simplest way of usage-based learning would be referring to example sentences while learning with a vocabulary book, researchers have investigated how computers can augment this learning strategy by leveraging big data on the Web [49, 90]. Specifically, existing online videos [90] or news articles [49] are used as a source of providing practical usages of the words to learn.

Considering that the effectiveness of these techniques for achieving micro or usage-based learning has been confirmed, achieving both simultaneously would be a promising way for further improving the experience of vocabulary learning using computers. However, how to combine both strategies effectively is not trivial; simply presenting the usages of the words to learn within a user's daily life would not be optimal because the usages available in existing resources (e.g., videos or news articles) do not always match the user's context. Thus, it is hard to offer the user contextual clues to learn the words so that the user can fully benefit from the strategy of micro learning.

Here, we suppose that the recent advance of natural language processing (NLP) techniques opens up a way to address this point; we can generate such usages by taking a user's context into account instead of retrieving them from existing resources. Then, the user can be exposed to contextualized usages of the words they want to remember in their daily lives without spending dedicated time studying. An example scenario is as follows:

> One night, Satoru, a Japanese student studying English words for a university exam, studied "bogus" in a vocabulary book. The next day, he surfed the Web and began reading some news articles in Japanese as usual. While reading, he encountered the following sentence that was partially translated into English: "ロックダウンの延長に伴う政府の給付金ですが[1]、*its applicants were mostly occupied by bogus companies.* これを受けて政府は...[2]" Here, he tried to recall the meaning of "bogus," given the context of the news article, and he eventually relearned its meaning with the presented example.

In this scenario, Satoru could easily memorize the meaning of a new English word he wanted to learn by encountering its usage while casually reading Japanese content on the Web. In a nutshell, our key idea is to *encapsulate* foreign words in materials that a user reads in their native language to expose them to contextualized usages of the words. While this assumes that the user has the least command of foreign language so that they can understand the presented usage, the user can learn the encapsulated words in their daily lives.

For this purpose, *VocabEncounter* is proposed by leveraging recent NLP techniques to generate such usages of foreign words on the spot. Specifically, our system first identifies phrases in the original materials that could be used to encapsulate one of the given foreign words that the user wants to remember. Here, multilingual word embedding [47] and dependency structure analysis [89] are utilized in combination. By doing so, we can extract phrases likely to preserve their original meaning and maintain syntactic naturalness when used for encapsulation. Secondly, the targeted phrases are translated into the foreign language so that the translated phrases include one of the given foreign words. This translation process is enabled by introducing a constrained decoding algorithm [38] into a Transformer-based neural machine translation (NMT) model [84]. We also introduce a scoring algorithm based on Sentence-BERT [70] to ensure that the translated phrases preserve the original meaning and maintain syntactic naturalness. Lastly, as a proof-of-concept, we develop an interface incorporating this encapsulation approach for vocabulary learning in everyday life in the form of a Chrome extension. The extension encapsulates the words to learn into Web content the user reads in real time by presenting the translated phrases along with a hoverbox for looking up word meanings if the user does not recall them.

Towards deploying VocabEncounter for vocabulary learning, there would be several challenges to overcome. For example, the translation mechanism may present unnatural or mistranslated phrases. Also, even if we can generate natural usages, experimental verification is required to show that encountering such usages helps users memorize foreign words. Furthermore, since VocabEncounter is intended to enable users to learn vocabulary during their daily lives without dedicating time, it should be tested whether the experience of learning with VocabEncounter is favored.

We conducted a series of experiments to demonstrate that our system sufficiently resolves these concerns. First, we examined the quality of the phrases generated by VocabEncounter using a crowdsourcing service. The result confirmed not only the sufficiency of the targeting mechanism but also the feasible quality of the translation mechanism. In particular, the translated phrases produced by the proposed approach received an evaluation comparable to those translated by a bilingual speaker under the same constraint of encapsulating specified words. Next, we conducted a user study where we asked participants to spend a day using VocabEncounter, which is implemented as a Chrome extension, on their PCs. The result suggested that not only the experience of being presented with foreign words to learn during Web browsing but also the design of presenting them with their generated usages help the participants memorize the words. Lastly, we conducted a one-week user study where we provided VocabEncounter to learners and let them use it freely. Through semi-structured interviews, we confirmed that they favored the experience of learning with VocabEncounter, especially due to its design of achieving micro and usage-based learning simultaneously. We want to emphasize that our proposed approach can be employed in various situations, as illustrated in Figure 1.

[2] Translation: "ロックダウンの延長に伴う政府の給付金ですが" ⟶ "With regard to the government benefits due to the lockdown extension"

[2] Translation: "これを受けて政府は..." ⟶ "In response to this, the government ..."

These prototypes, along with the above results, exhibit that our presented paradigm involving NLP-based generation techniques can transform our daily activities into a field for vocabulary learning.

2 RELATED WORK

To situate our work, we start by reviewing existing interaction techniques for vocabulary learning. As mentioned in Section 1, there are two main strategies in supporting vocabulary learning: micro and usage-based learning. In this section, we discuss how previous HCI works have leveraged computers for each learning strategy along with the pedagogical background behind them. We then introduce related NLP techniques that enable us to achieve the scenario envisioned in Section 1.

2.1 Interaction techniques for vocabulary learning: micro learning

While the importance of vocabulary learning in mastering a foreign language has been acknowledged for a long time [1, 4], most learners have relied on simple educational resources like word lists or vocabulary cards [61]. To help them learn vocabulary more effectively, a large body of HCI research has been devoted to introducing various interaction techniques. One major objective is delivering *micro* learning [30, 40, 45] with the help of mobile or Web technologies.

In micro learning, learners are encouraged to leverage small learning units and short-term learning activities within their daily lives. This learning strategy was introduced to address the difficulty of learning vocabulary in the busyness of everyday life [30, 40]. While learners need to be constantly exposed to the words they want to remember to overcome the forgetting curve [67], we often fail in scheduling dedicated time for studying, which motivates the exploration of practical ways to learn in small time segments. This learning strategy can also be associated with the concept of casual learning [34, 44] in terms of its emphasis on leveraging daily opportunities to decrease the mental burden of studying. By incorporating this strategy, learners can reduce their cognitive load while maintaining long-term retention [50].

A wide range of computing devices has been exploited to achieve micro learning [14, 15, 20, 22, 23, 25]. For example, Dingler et al. [22] explored the use of pervasive physical displays for vocabulary learning by placing displays throughout users' homes and work environments. Smartphones are often leveraged to pursue more handy approaches, such as presenting quizzes through notification [23] and scheduling tests repeatedly [25]. Some work also made use of small time segments associated with smartphones, such as using live wallpaper [20] and wait time in messaging and loading [14, 15].

In addition to leveraging small time segments, some recent techniques exploit the contextual information of learners because grounding new words in learners' context is known to be effective for enhancing their memorization [33, 58]. For example, RFID-based activity recognition [8, 64] and GPS-based location recognition [26, 36] were used to present foreign words or vocabulary quizzes of relevance to a user's context. Smart Web browsers for vocabulary learning [10, 81] have been proposed to capture the context while using computers. Berleant et al. [10] proposed an approach of translating words on a Web page into a foreign language to increase users' exposure to the words. The effectiveness of such a word-level translation approach on the user's memory performance has been empirically confirmed by Trusty and Truong [81]. These studies imply the importance of transforming our daily lives into a field of vocabulary learning. At the same time, as computers have increased their presence in our lives, they also signify the potential of HCI techniques for facilitating micro learning.

2.2 Interaction techniques for vocabulary learning: usage-based learning

The other objective is supporting vocabulary learning via *usage-based* learning. In fact, most conventional educational resources present not only the formal definitions of words but also their usages in the form of example sentences [60]. Example sentences are offered because, without seeing usages, it is difficult to elaborate the semantic information of words, which is crucial to the long-term retention of the words [13]. In particular, it is experimentally confirmed that even example phrases consisting of as few as ten words were effective in acquiring vocabulary [6].

Given this context, a new paradigm of *usage-based* learning has been recently introduced by leveraging big data on the Web [49, 78, 79, 90], allowing users to learn the practical usages of words they want to remember by automatically retrieving them from the Web. Syed and Collins-Thompson [78, 79] have proposed ways to tailor web search rankings to support vocabulary learning where pages that contain words for a learner to remember are prioritized. Lungu et al. [49] proposed a personalized system that recommends Web articles that are likely to contain words unknown to each user. Moreover, Vivo is a video-augmented dictionary that provides a way to exploit huge online video resources for vocabulary learning [90]. It appropriately identifies short movie scenes that contain the usages of the words a user wants to remember from existing movies and provides the user with them as a contextual clue.

Although using these systems positively affects vocabulary learning, they still rely on the user's motivation to watch or read the recommended contents in a foreign language, which are not always attractive to every user. This point conversely suggests that combining them with interaction techniques for supporting micro learning [30, 40] can deliver vocabulary learning more effectively. However, how to achieve the combination is not trivial because, even with big data, it is not always possible to get the usages of the words that users want to remember in a way that is consistent with their context. In fact, the smart Web browser implemented by Trusty and Truong [81] limited the foreign words to be presented while browsing to those included in their hand-crafted dictionary of 1,500 nouns. They rationalized it by stating "verbs and other parts of speech are highly dependent on context [81]," which implies that their browser of simply replacing a single word on each Web page with its translation would not be optimal for presenting words in a contextualized manner. This taught us that achieving both micro and usage-based learning effectively requires us to overcome the difficulty of presenting the usages in a manner that is grounded in the user's context.

Here, we suppose that the recent advance of NLP techniques enables us to overcome this difficulty; i.e., we can generate the usages of words a user wants to remember in a contextualized manner. By *encapsulating* the words with their generated usages into materials the user is exposed to in their daily lives, such as during Web browsing, the user can learn them while casually reading the materials in their native language, without taking dedicated time. This approach is expected to be beneficial given the finding by Brooks [12] that learners positively perceive the experience of actively incorporating their first language in foreign language acquisition, compared to focusing only on the language to learn. Then, our main challenge is how to achieve such encapsulation naturally while maximizing the opportunities for users to encounter the words. Our approach tackles this by leveraging recent NLP techniques, as described in Section 2.3.

2.3 Related NLP techniques

To build VocanEncounter as an automated system, we employed three NLP techniques: NMT with constrained decoding, multilingual word embedding, and sentence embedding. In this section, we would like to introduce them to provide background for our approach.

Machine translation systems have been significantly improved by deep neural networks [5, 42, 76, 84]. The output sentences are high quality; in some domains, they can be comparable to professional human translators, i.e., one cannot tell human translations from system translations [7, 35]. As well as the quality, controllability is also studied, which leads to the development of constrained decoding [3, 37, 38, 68]. It enables us to generate translations with target words specified dynamically at the time of translation, while other methods often require retraining of the NMT models with the required words [24, 75]. In this study, we incorporated the approach proposed by Hu et al. [38] with a good trade-off between accuracy and speed for near real-time translation.

Word embedding has also become the de facto standard approach for obtaining the numerical representation of words and efficiently calculating their semantic similarity [9, 16, 19, 53]. Afterward, some advanced work enables embedding words from multiple languages into the joint semantic vector space, where similar words can be found cross-lingually (e.g., car [En], wagen [De], voiture [Fr], 車 [Ja]) [52, 72]. We can obtain such a semantic space even in unsupervised manners, that is, without multilingual word dictionaries [47].

The idea of this semantic embedding has also been extended to the sentence level. While an empirical method of averaging the embedding vector of all words in a sentence has been widely used, dedicated methods for mapping variable-length sentences to a semantic vector space have also evolved [17, 43, 70]. For example, Sentence-BERT [70] leverages BERT [21], a pretrained Transformer-based [84] model, to derive semantic embedding vectors that can be compared using cosine similarity. In an analogous manner to word embedding, its multilingual extension has also been proposed [27, 71].

These techniques allow us to construct a novel vocabulary-learning system that achieves the two learning strategies discussed in Section 2.1 and Section 2.2. In particular, the human-level automated translation techniques enable a novel approach of supporting vocabulary learning that dynamically generates the usages of the words to learn, rather than retrieving them from the Web. By incorporating such NMT techniques with semantic embedding vectors that can be calculated in near real-time, the proposed system can encapsulate the words a user wants to learn into materials the user is reading by taking the user's context into account.

3 PROPOSED APPROACH

In this section, we first explain how *VocabEncounter* is designed to achieve both micro and usage-based vocabulary learning taking a user's context into account. We then describe how VocabEncounter encapsulates the words a user wants to learn through its two core mechanisms: targeting and translating. We also explain how the usages generated by VocabEncounter are presented in the user's daily activities by showing our user interface implemented as a Chrome extension.

3.1 Overview

Our aim is to support vocabulary learning through VocabEncounter by presenting the contextualized usage of the words that users want to learn during their daily lives (Figure 1). This is enabled by leveraging recent NLP techniques to automatically generate translated phrases containing specified foreign words in near real time. As proof of concept, we considered utilizing the user's everyday browsing experience, inspired by previous works suggesting its efficacy as micro learning [10, 81]. In this case, the entire encapsulation process is illustrated in Figure 2.

Looking back to the example scenario in Section 1, let us consider a user who is a native Japanese speaker and studying English. We note that our system can be used with other pairs of languages as long as there is sufficient data to train the models we used. As we intend to offer a personalized learning experience, we assume a specific set of foreign words $W^{En} = \left\{w_1^{En}, w_2^{En}, \cdots, w_j^{En}, \cdots, w_N^{En}\right\}$ that the user wants to learn. Typically, W^{En} can be given from word lists or determined by a vocabulary test (i.e., words the user could not identify their correct meaning). Then, VocabEncounter provides the user with the opportunity to encounter the words by presenting translated phrases containing those words in documents $D^{Ja} = \left\{d_1^{Ja}, d_2^{Ja}, \cdots, d_i^{Ja}, \cdots, d_M^{Ja}\right\}$. Here, d_i^{Ja} denotes a document retrieved from each page the user visits during their Web browsing.

The process of presenting the translated phrases runs whenever the user visits a Web page. VocabEncounter first targets phrases to be used for the encapsulation from the page. This process is designed to prioritize phrases whose English translations are supposed to preserve their original meaning and maintain syntactic naturalness upon encapsulation. Formally, it enumerates pairs of a word in W^{En} and a phrase in d_i^{Ja} like $\left\{\left(w_1^{En}, p_{i,1}^{Ja}\right), \left(w_3^{En}, p_{i,3}^{Ja}\right), \cdots, \left(w_N^{En}, p_{i,N}^{Ja}\right)\right\}$ where $p_{i,j}^{Ja}$ is a phrase in d_i^{Ja} to be translated so as to contain w_j^{En}. Note that there are words in W^{En} that may not appear on the page if there is no appropriate phrase in d_i^{Ja}, as w_2^{En} in this case (Figure 2).

VocabEncounter then translates each $p_{i,j}^{Ja}$ into an English phrase $p_{i,j}^{En}$ under the constraint of containing w_j^{En} and scores the degree

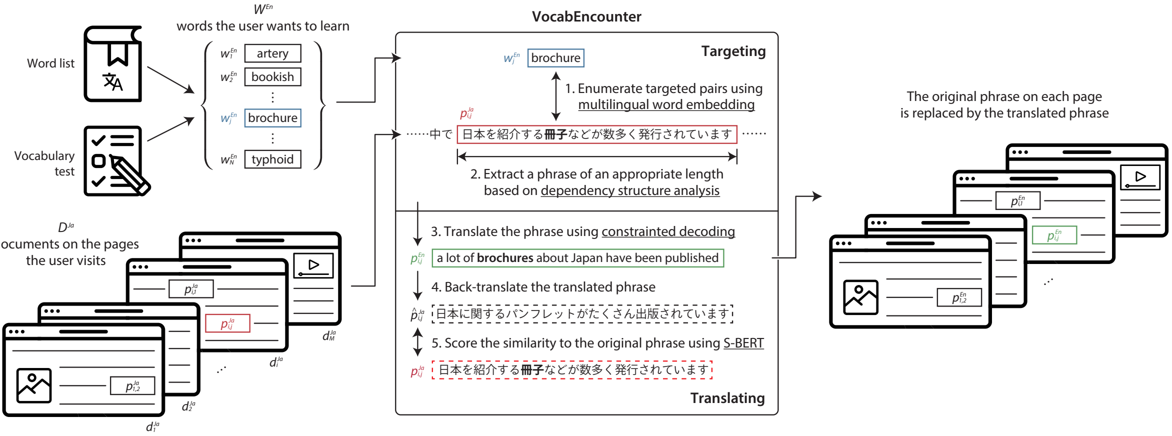

W^{En}
words the user wants to learn
Word list
Vocabulary test
w_1^{En} artery
w_2^{En} bookish
w_j^{En} brochure
w_N^{En} typhoid
D^{Ja}
documents on the pages the user visits
$p_{i,1}^{Ja}$
$p_{i,j}^{Ja}$
$p_{1,2}^{Ja}$
d_1^{Ja}
d_2^{Ja}
d_i^{Ja}
d_M^{Ja}
VocabEncounter
Targeting
w_j^{En} brochure
1. Enumerate targeted pairs using multilingual word embedding
$p_{i,j}^{Ja}$
……中で 日本を紹介する冊子などが数多く発行されています ……
2. Extract a phrase of an appropriate length based on dependency structure analysis
3. Translate the phrase using constrainted decoding
$p_{i,j}^{En}$ a lot of brochures about Japan have been published
4. Back-translate the translated phrase
$\hat{p}_{i,j}^{Ja}$ 日本に関するパンフレットがたくさん出版されています
5. Score the similarity to the original phrase using S-BERT
$p_{i,j}^{Ja}$ 日本を紹介する冊子などが数多く発行されています
Translating
The original phrase on each page is replaced by the translated phrase
$p_{i,j}^{En}$
$p_{1,2}^{En}$

Figure 2: Example pipeline of how VocabEncounter presents translated phrases containing specified foreign words in our proof-of-concept Chrome extension.

of the translated phrase $p_{i,j}^{En}$ having the meaning of $p_{i,j}^{Ja}$. If $p_{i,j}^{En}$ is judged to sufficiently preserve the original meaning, our system substitutes $p_{i,j}^{Ja}$ on the page with $p_{i,j}^{En}$. In addition, $p_{i,j}^{Ja}$ is presented in a hoverbox along with the meaning of the word w_j^{En} in Japanese when the user places a mouse pointer over $p_{i,j}^{En}$, as we exhibit later in Figure 4.

3.2 Targeting

As mentioned in Section 3.1, VocabEncounter first performs the targeting process to identify Japanese phrases on the page to encapsulate W^{En}. Without this process, it needs to translate a lot of combinations across all phrases on the page and all words in W^{En}, which takes a long time and makes it impossible to present translated phrases while the user is on the page. Instead, our system filters out combinations whose translation results would not preserve the original meaning nor maintain syntactic naturalness.

More specifically, for each w_j^{En}, VocabEncounter identifies Japanese words in d_i^{Ja} that have a similar meaning to w_j^{En}. This is based on the assumption that phrases containing such words would be suitable for encapsulating w_j^{En} in terms of preserving the original meaning and maintaining syntactic naturalness. The process is enabled using multilingual word embedding [47], trained to map similar words onto close embedding vectors regardless of the language. Thus, we can measure the similarity between w_j^{En} and each Japanese word in d_i^{Ja} by calculating the cosine similarity of their corresponding embedding vectors. Then, our system lists Japanese words whose similarity exceeds a threshold th_1.

For each of the listed words, VocabEncounter first extracts a Japanese sentence containing the word from d_i^{Ja}. Here, it does not translate the whole sentence into English but further extracts a phrase of an appropriate length containing the word from the sentence. This is because sentences on the Web are sometimes too long, yielding long translation results as well. On the other hand, it was confirmed that example sentences of about ten words are effective in vocabulary learning [6]. Thus, we designed VocabEncounter not to present long translation results that would unnecessarily consume the user's cognitive load but to extract a phrase of a certain length (e.g., 10 or 20 words) before translating.

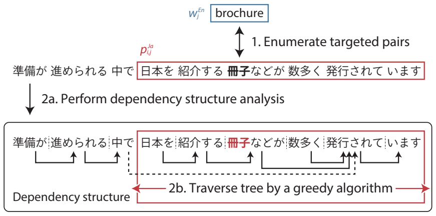

w_j^{En} brochure
1. Enumerate targeted pairs
$p_{i,j}^{Ja}$
準備が 進められる 中で 日本を 紹介する 冊子などが 数多く 発行されて います
2a. Perform dependency structure analysis
準備が 進められる 中で 日本を 紹介する 冊子などが 数多く 発行されて います
Dependency structure
2b. Traverse tree by a greedy algorithm

Figure 3: Example pipeline of how VocabEncounter extracts a phrase of a certain length to translate given the pair of d_i^{Ja} and w_j^{En}.

For this purpose, VocabEncounter applies dependency structure analysis [89] (specifically, ginza[3] trained on [65]) to the sentence to obtain its structure tree. Here, naïve algorithms such as extracting a certain number of words before and after the targeted word can produce unnatural chunks that are hard to translate. Instead, our system extracts a syntactically compositional phrase based on the structure tree using a greedy algorithm.

Specifically, let us consider a case in which a word "brochure" is identified to have a similar meaning to a word "冊子", as illustrated in Figure 3. VocabEncounter traverses the dependency structure tree starting from the shortest phrase containing the targeted word (i.e., "冊子"). It then tries to expand the phrase by concatenating precedent or subsequent words, prioritizing ones posing a smaller number of dependencies that refer to words outside the extracted phrase. The expansion is repeated until the number of words in the phrase to be extracted reaches 10.

[3] https://megagonlabs.github.io/ginza/

3.3 Translating

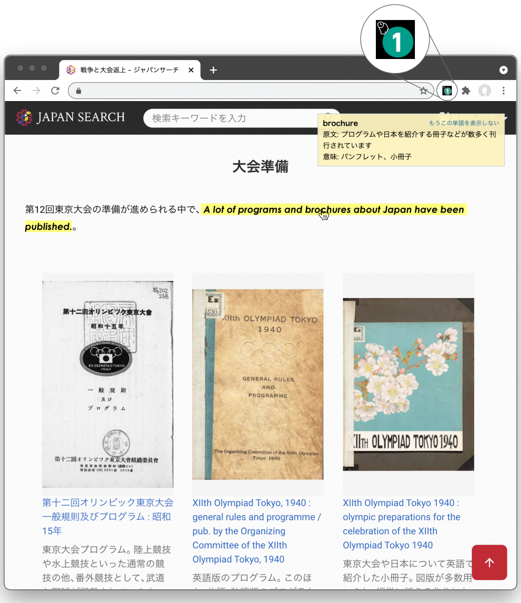
Now, VocabEncounter has pairs of an English word w_j^{En} in W^{En} and a Japanese phrase $p_{i,j}^{Ja}$ in d_i^{Ja}. We want to translate $p_{i,j}^{Ja}$ into an English phrase $p_{i,j}^{En}$ so as to contain w_j^{En} while preserving the original meaning of $p_{i,j}^{Ja}$ and maintaining syntactic naturalness. Here, a vanilla translation using an NMT model would not satisfy the requirement of containing the specified word. Thus, our system incorporates the NLP technique described in Section 2.3, specifically, a constrained decoding algorithm proposed by Hu et al. [38]. Their algorithm can easily switch the word to be contained at no extra cost like retraining the model, as mentioned in Section 2.3. This is desirable for use in VocabEncounter, where it is possible that the user updates W^{En} in the short term; for example, the user may take a vocabulary test every day to remove learned words and add new unknown words to W^{En}.

Before presenting the translated phrase $p_{i,j}^{En}$, VocabEncounter judges whether $p_{i,j}^{En}$ sufficiently preserves the original meaning of $p_{i,j}^{Ja}$. We expected this to contribute to the user's experience by filtering out translated phrases that are not much like the original phrases. For this purpose, it performs a backward translation from $p_{i,j}^{En}$ to $\hat{p}_{i,j}^{Ja}$ and measures the similarity between the original phrase $p_{i,j}^{Ja}$ and the back-translated phrase $\hat{p}_{i,j}^{Ja}$. The similarity between the two Japanese phrases is obtained by calculating their semantic embedding vectors using Sentence-BERT [70] and measuring the cosine similarity of the vectors. Here, we employed this backward translation approach rather than multilingual Sentence-BERT [27, 71] because the similarity through a round-trip translation is known to reflect the quality of the forward translation [55, 63]. Thus, this approach would also contribute to ensuring the naturalness of $p_{i,j}^{En}$.

Still, VocabEncounter determines whether to present the translated phrase not only by relying on the similarity score. Alternatively, in the same way that standard NMT techniques [84] determine the best translation result from candidates based on their likelihood score, we also considered the likelihood score of the phrase evaluated by the translation model. Specifically, a higher likelihood score reflects that the phrase is more likely to appear in the training data of the translation model, giving more weight to natural phrases. We considered this likelihood score because, in our empirical observations, using only the similarity score sometimes allowed presenting less syntactically natural phrases. We found that balancing the similarity score and the likelihood score, as follows, would be suitable for maintaining syntactic naturalness while preserving the original meaning.

$$Score\left(p_{i,j}^{Ja}, p_{i,j}^{En}, \hat{p}_{i,j}^{Ja}\right) = \frac{2\,\textsc{SBert}\left(p_{i,j}^{Ja}, \hat{p}_{i,j}^{Ja}\right) + \mathrm{L}_{Ja \mapsto En}\left(p_{i,j}^{Ja}, p_{i,j}^{En}\right) + \mathrm{L}_{En \mapsto Ja}\left(p_{i,j}^{En}, \hat{p}_{i,j}^{Ja}\right)}{4} \tag{1}$$

Here, $\textsc{SBert}(\cdot)$ denotes the similarity score obtained by Sentence-BERT [70], and $\mathrm{L}(\cdot)$ denotes the likelihood score obtained by the translation model. Then, our system presents $p_{i,j}^{En}$ to the user if $Score\left(p_{i,j}^{Ja}, p_{i,j}^{En}, \hat{p}_{i,j}^{Ja}\right)$, ranging from 0 to 1, exceeds a threshold th_2.



戦争と大会返上 - ジャパンサーチ

JAPAN SEARCH

検索キーワードを入力

brochure もうこの単語を表示しない

原文: プログラムや日本を紹介する冊子などが数多く刊行されています

意味: パンフレット、小冊子

大会準備

第12回東京大会の準備が進められる中で、*A lot of programs and brochures about Japan have been published.*。

第十二回オリンピック東京大会一般規則及びプログラム : 昭和15年

東京大会プログラム。陸上競技や水上競技といった通常の競技の他、番外競技として、武道

XIIth Olympiad Tokyo, 1940 : general rules and programme / pub. by the Organizing Committee of the XIIth Olympiad Tokyo, 1940

英語版のプログラム。このほ

XIIth Olympiad Tokyo 1940 : olympic preparations for the celebration of the XIIth Olympiad Tokyo 1940

東京大会や日本について英語で紹介した小冊子。図版が多数用

1
第十二回オリンピツク東京大會
昭和十五年
一般規則
及び
プログラム
第十二回オリンピツク東京大會組織委員會
XIIth OLYMPIAD TOKYO
1940
GENERAL RULES
AND
PROGRAMME
The Organizing Committee of the XIIth Olympiad Tokyo 1940
XIITH OLYMPIAD TOKYO 1940

Figure 4: VocabEncounter presents a translated phrase containing specified words and supplies its original phrase with its meaning when the user places a mouse pointer over the translation[5]. The number of translated phrases presented on the current page is indicated next to the address bar to encourage users to be aware of them.

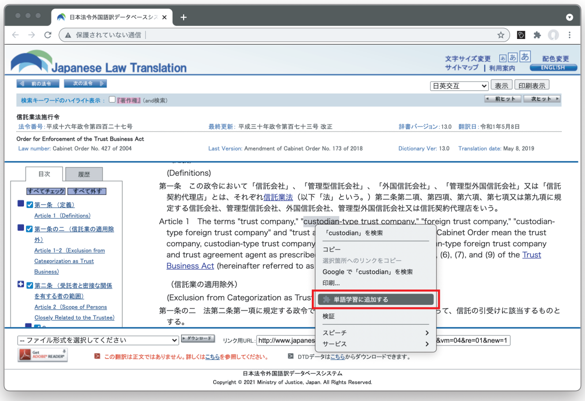
We implemented this translation process using fairseq [66] with English–Japanese and Japanese–English translation models pretrained on JParaCrawl [56], a large-scale dataset made from various resources on the Web. In addition, as we mentioned above, the translated phrases should be presented before the user leaves the page. To satisfy this, we designed to execute the translation on a GPU-powered server[4]by communicating from a browser through Websocket.

3.4 User interface

In our proof-of-concept implementation, the above processes are delivered to the user in the form of a Chrome extension. When the user visits a new Web page, the extension first retrieves all text nodes on the page and sends them to the server via Websocket. Then, if the server finds appropriate phrases to encapsulate W^{En} and returns pairs of $p_{i,j}^{En}$ and $p_{i,j}^{Ja}$, the extension replaces the original phrase $p_{i,j}^{Ja}$ with $p_{i,j}^{En}$. The extension also provides a hoverbox presenting the meaning in Japanese to help the user if they do not recall the meaning of the encapsulated word w_j^{En}, which is also sent from the server along with the translated results. When the user places a mouse pointer on $p_{i,j}^{En}$, the hoverbox appears, as presented in Figure 4.

[4] Specifically, we used a server with a NVIDIA Geforce GTX 1080 Ti.

[5] The content on this page is allowed to use or modify under the CC BY 4.0 by National Diet Library, Japan.



日本法令外国語訳データベースシス

保護されていない通信

Japanese Law Translation

信託業法施行令

法令番号：平成十六年政令第四百二十七号　最終更新：平成三十年政令第百七十三号 改正　辞書バージョン：13.0　翻訳日：令和1年5月8日

Order for Enforcement of the Trust Business Act

Law number: Cabinet Order No. 427 of 2004　Last Version: Amendment of Cabinet Order No. 173 of 2018　Dictionary Ver: 13.0　Translation date: May 8, 2019

(Definitions)

第一条　この政令において「信託会社」、「管理型信託会社」、「外国信託会社」、「管理型外国信託会社」又は「信託契約代理店」とは、それぞれ信託業法（以下「法」という。）第二条第二項、第四項、第六項、第七項又は第九項に規定する信託会社、管理型信託会社、外国信託会社、管理型外国信託会社又は信託契約代理店をいう。

Article 1　The terms "trust company," "custodian-type trust company," "foreign trust company," "custodian-type foreign trust company" and "trust a Cabinet Order mean the trust company, custodian-type trust company n-type foreign trust company and trust agreement agent as prescribe , (6), (7), and (9) of the Trust Business Act (hereinafter referred to as

（信託業の適用除外）

(Exclusion from Categorization as Trus

第一条の二　法第二条第一項に規定する政令で って、信託の引受けに該当するものとする。

この翻訳は正文ではありません。詳しくはこちらを参照してください。

DTDデータはこちらからダウンロードできます。

日本法令外国語訳データベースシステム

Copyright © 2021 Ministry of Justice, Japan. All Rights Reserved.

文字サイズ変更
あ あ あ
配色変更
サイトマップ | 利用案内
ENGLISH
前の法令
次の法令
日英交互
表示
印刷表示
検索キーワードのハイライト表示：□『著作権』（and検索）
前ヒット
次ヒット
目次
履歴
すべてチェック
すべて外す
第一条（定義）
Article 1 (Definitions)
第一条の二（信託業の適用除外）
Article 1-2 (Exclusion from Categorization as Trust Business)
第二条（受託者と密接な関係を有する者の範囲）
Article 2 (Scope of Persons Closely Related to the Trustee)
「custodian」を検索
コピー
選択箇所へのリンクをコピー
Google で「custodian」を検索
印刷...
単語学習に追加する
検証
スピーチ
サービス
-- ファイル形式を選択してください
ダウンロード
リンク用URL：http://www.japanes
&vm=04&re=01&new=1
Get ADOBE READER

Figure 5: VocabEncounter allows the user to easily add a word to be used for encapsulation when they find an unknown word on a page and want to remember by right-clicking the word on a Web page[6].

The user can freely add a new word to W^{En} on the options page of the extension or just by right-clicking a foreign word on a Web page that the user does not know, as shown in Figure 5. In contrast, if the user remembers one of the words in W^{En} through being familiarized with its usages they encountered, the user can easily remove the word from W^{En} on the options page or by indicating it by clicking a button that is placed in the upper right corner of the hoverbox (Figure 4).

To facilitate the user's experience of vocabulary learning with this extension, VocabEncounter employs an incremental translation strategy. It translates the targeted phrases in the order of their appearance on the page and returns the translation results incrementally before all phrases are translated. By doing so, phrases on the beginning of the page would be translated immediately, and phrases later in the page would be processed while the user reads the page from the beginning. In addition, the translation process of each page is queued in a LIFO (last-in-first-out) order so that the user can see the translation results of the current page even while the translation process of the previous pages is ongoing. Furthermore, to maximize opportunities of encountering the words to learn, our system indicates the number of translated phrases on the current page, as shown in Figure 4, encouraging the user to be aware of the translated phrases.

4 PILOT TEST

Before implementing the full version of VocabEncounter, we first tested the feasibility of the proposed approach by implementing only the targeting mechanism (Section 3.2). This is because it was possible that the targeting mechanism may not sufficiently identify phrases to encapsulate the words to memorize, which hinders users from encountering the words. Alternatively, it may enumerate too many phrases, making it impossible to translate all of them in near real time while the users are on the page. In this sense, we considered a day-by-day browsing experience as in the example scenario in Section 1 and examined how many words of W^{En} would go to the translation process after the targeting process is applied to documents that an average user browses in a day.

For this purpose, we first gathered volunteers and constructed a dataset of D^{Ja} by collecting documents from the pages they visited in a day. We also prepared a public English word list, assuming that an average learner would set W^{En} using the list based on their English level. In terms of simulating the learner's behavior of using VocabEncounter for a day, it is possible to set W^{En} by randomly sampling 10 or 20 words and examine how many of them would be encapsulated. However, to assure the generality, we virtually considered the case of specifying all words in the list (about 2,500 words) as W^{En}. Then, we applied the targeting mechanism (Section 3.2) and evaluated how many of the words in the list could be used for the encapsulation.

4.1 Data collection

To construct a dataset of D^{Ja} reflecting the browsing experiences of average users, we gathered volunteers through word-of-mouth and online communication. Each volunteer was asked to install a recording tool in their browser that collected documents on all the pages they visited. Before installing the tool, an experimenter from the authors carefully revealed its function and informed them that their data would be anonymized for analysis. We also recommended that they switch to a different account or browser when they are accessing data that is inappropriate to send, for example, reading confidential documents. When they agreed to use the tool, the collection started and lasted for a day. The exact timing of start and end was left to each volunteer. After all, five volunteers participated, resulting in a dataset of five D^{Ja}.

To decide on the list of English words, we referred to the Common European Framework of Reference for Languages (CEFR) [54], an international standard for classifying language ability. It describes language ability on a six-point scale, from A1 for beginners up to C2 for those who have mastered the language. We chose B2, a semi-advanced level used for describing those confident in using the language because it is slightly higher than the average level for Japanese (B1) [29]. Therefore, we expected that the corresponding vocabulary list would contain some unknown words for most people, matching the actual study needs. As a result, we adopted all 2,692 English words that correspond to B2 from a public word list validated with more than 5,000 students [80].

4.2 Analysis

As we discussed in Section 3.2, the function of the targeting mechanism highly depends on its threshold (th_1). A higher value of th_1 filters out a greater number of the associated pairs of words and phrases, reducing the chances that each word of W^{En} would be encapsulated in D^{Ja}. In contrast, lower values of th_1 increase the number of phrases to be translated, taking more time before they are presented to users. Thus, we applied the targeting mechanism with different th_1 using the collected dataset of D^{Ja} and the chosen list of English words and examined this point.

[6] The content on this page is allowed to quote, reproduce, and reprint by Ministry of Justice, Japan.

Specifically, we first calculated the ratio of words that never matched with any phrases in the dataset. Such cases are undesirable as users would not get an opportunity to encounter the word, failing to learn it. For each D^{Ja} in the dataset, we applied the targeting mechanism of VocabEncounter with a certain th_1 under the assumption that all words in the list were specified as W^{En}. We then counted the number of words that were never used in the extracted phrases and calculated the ratio of that number to all words in the list. Finally, we averaged the ratio over a set of D^{Ja}. We repeated this process with different th_1 and checked the averaged ratio.

As a result, we determined that, when the threshold for targeting (i.e., th_1) was set to be 0.5, more than 80% of the words, on average, were used for the encapsulation at least once. From this result, we can expect users to be exposed to most of the words they want to remember within a day. In addition, we empirically confirmed that, when we set $th_1 = 0.5$ with W^{En} of around 30 words, the targeting mechanism enumerates at most about 10 pairs of a word and a phrase on regular Web pages, which can be translated in near real time while users are on the pages. These points assured us that VocabEncounter could be implemented to allow users to encounter the words to learn during Web browsing. We thus implemented the full version of VocabEncounter (i.e., with the translation mechanism and user interface) and evaluated its effectiveness in the following sections.

5 HYPOTHESES

Up to this point, we have introduced VocabEncounter as a system for vocabulary learning, which can present the usages of the words to learn during daily browsing in a contextualized manner. The envisioned scenario and design of VocabEncounter, which we discussed in Section 1 and Section 3.1, respectively, impose several hypotheses. These hypotheses are required to be verified to ascertain the feasibility and effectiveness of VocabEncounter in transforming our daily life into a field of vocabulary learning.

First, it should be examined whether VocabEncounter can generate translated phrases containing specified words with quality enough to support users learning vocabulary during their browsing. Specifically, we need to present natural word usages to familiarize users with how the words are used. Moreover, the translated phrases should preserve the meaning of their original phrases and not interfere with users reading the content on a Web page. Thus, the following H1 is posited.

> H1: VocabEncounter can generate natural usages of specified foreign words by translating phrases on the Web without losing their original meaning.

If H1 holds, it then means that VocabEncounter is capable of offering usage-based learning during browsing experiences. As discussed in Section 2.2, this would allow the users to fully benefit from both micro and usage-based learning through being familiarized with the contextualized usages of the words to learn in their daily lives.

We then expect that VocabEncounter can lead to a better learning outcome than simply presenting the words to learn in the same manner as conventional micro learning approaches (Section 2.1). Thus, our second hypothesis is the following:

> H2: VocabEncounter can induce a better learning outcome than simply showing the words to remember during Web browsing through presenting their generated usages.

Lastly, in order for VocabEncounter to be practically adopted for supporting vocabulary learning, its usability and efficacy when used in users' daily lives should be investigated. As envisioned in the example scenario in Section 1, we anticipate that users would favor the minimized hurdle for studying delivered by the experience of encountering the usages of the words they want to remember during daily browsing, positing the following hypothesis:

> H3: The experience of learning with VocabEncounter is favored by users when used in their daily lives, thanks to its design of achieving both micro and usage-based learning.

If these hypotheses are supported, we can conclude that VocabEncounter opens up a new learning strategy that embraces the advantages of both micro and usage-based learning through an NLP-powered approach, which can transform our daily lives into a field of vocabulary learning. With this motivation, we evaluated these hypotheses by conducting a series of experiments.

6 EXPERIMENT I: QUALITY OF THE TRANSLATED PHRASES

We first conducted an offline experiment to evaluate the quality of phrases translated by VocabEncounter under the constraint of containing one of the specified words in correspondence of H1. Specifically, we asked human assessors fluent in Japanese and English to check whether the translated phrases are natural while preserving the original meaning. As a baseline for comparison, we prepared two conditions: phrases translated by a human translator with the same constraint of using the specified words and phrases translated by the same NMT model without the constraint of using the specified words.

6.1 Design and material

We first introduce the *proposed* condition, which refers to the output of VocabEncounter; i.e., the phrases translated by the NMT model with the constrained decoding algorithm so as to contain the specified English words, as described in Section 3.3. The next one is the *human* condition, representing the phrases translated by a human translator who is bilingual and lived in the UK for years. The person voluntarily participated before being told about the concept of the experiment and was asked to translate the Japanese phrases while using the corresponding specified words. If the evaluations by human assessors were not significantly different between the *proposed* and *human* conditions, it is implied that our system achieves at least human-level quality under the constraint of using the specified word.

Furthermore, we prepared the *vanilla* condition, which is the output of the same pretrained NMT model with the *proposed* condition [56] but translated without the constraint of using the specified words. The *vanilla* condition was introduced as the best-effort case. By comparing its evaluations with those of the other two, we can know how difficult the translation task involving the encapsulation is and how well the proposed algorithm would work within the task.

To support H1, we expect that the *proposed* condition works comparably with the *human* condition, assuring that VocabEncounter can generate the usages with a certain level of naturalness and meaning preservation such that they can be used for vocabulary learning. Moreover, we expect that the phrases generated by the *proposed* condition, i.e., with the constraint of using the specified word in the translation, are not significantly different from those generated by the *vanilla* condition, i.e., without such a constraint. If these points are confirmed, users can encounter natural usages of the words contextualized in their reading content.

To prepare sample phrases for evaluation, we randomly extracted 60 pairs of English words and corresponding Japanese phrases targeted in the pilot test (Section 4). Then, we obtained three English phrases corresponding to the above three conditions for each of the extracted pairs, yielding 180 phrases in total. We present some of the phrases we obtained in this experiment (see Table 1 in Appendix).

6.2 Measure

To reflect H1, we examined the quality of the translated phrases in terms of their naturalness and meaning preservation. Both of these were measured through questionnaires using a 5-point Likert scale (with 1 indicating "strongly disagree" and 5 indicating "strongly agree"). We first showed a translated English phrase to assessors and asked them, "How natural is this phrase in English?" Then, we also showed its original Japanese phrase to them and asked, "How well does the English phrase preserve the meaning of the original Japanese phrase?"

6.3 Procedure

To compare the measures introduced in Section 6.2 across the three conditions, we used a crowdsourcing service and gathered 24 Japanese assessors. They self-reported a high level of proficiency in English which was equivalent to or higher than CEFR B2. Each assessor evaluated 20 phrases for each condition (60 phrases in total) using the questionnaires, as explained in Section 6.2. As illustrated in Figure 6, we divided the assessors into six groups of four to balance the assignment of the conditions. Then, the assessors evaluated the phrases one by one in random order. The entire process took approximately 45 minutes, and we paid approximately $10 as a reward to each assessor.

6.4 Results

According to the Kruskal–Wallis test ($p = 0.001$), we found a significant difference among the three conditions for the naturalness measure. We thus conducted a post-hoc test and confirmed that the *vanilla* condition obtained a significantly better evaluation than the others ($p < 0.05$), as presented in Figure 7. However, we could not find a significant difference between the evaluations for the *proposed* and *human* conditions. We further verified the equivalence of the two conditions using two one-sided tests based on the Mann–Whitney test [86] with an equivalence margin θ of 0.5, as Joosse et al. [41] did. As a result, the equivalence between the evaluations for the *proposed* and *human* conditions was supported.

The results indicate that, although the words to be encapsulated were selected through the targeting mechanism (Section 3.2), the constraint of containing the words can affect the naturalness of translation results. This can be confirmed by examining the translation results shown in #8 and #9 of Table 1 (from Appendix). In the former case, the word "pluck" was targeted in association with "引っ張る [pull]" and contained in the translation results of the *proposed* and *human* conditions. While "pluck" has a similar meaning to "pull," people would not use it frequently when they talk about muscle training. This subtle difference in nuance could reduce the naturalness of the translations. We also found an interesting phenomenon in the latter case, that is, the proposed method tried to contain the word "prom," which was associated with "デート [date]," but eventually contained the word "promise." This is due to the implementation of the constrained decoding algorithm, which uses a subword tokenizer [46] to allow the conjugation of the specified word by sometimes concatenating multiple (sub)words. While the obtained translation can convey the meaning of the original phrase, this case could contribute to the reduction of the naturalness and, in turn, suggests room for improvement in our implementation (see Section 11.2).

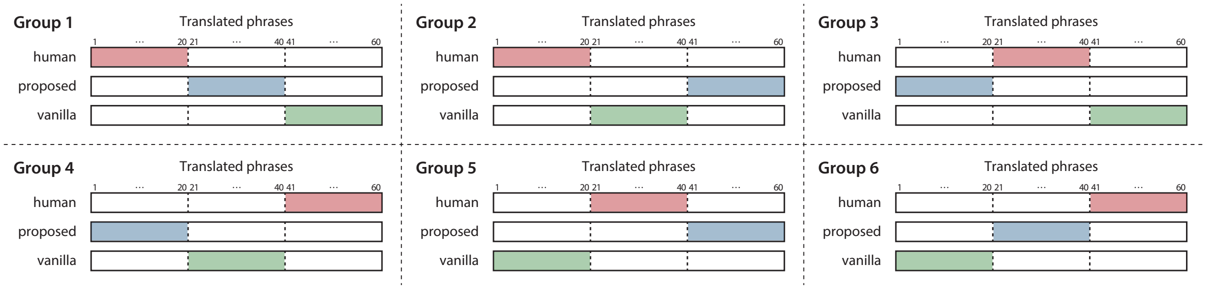
Yet, as we can infer from the fact that the effect size of the comparison between the *proposed* and *vanilla* conditions was relatively small ($r = 0.109$), it does not immediately imply the infeasibility of the proposed method in presenting the usages of the words. Notably, the naturalness of translation results given by the proposed method was at least equivalent to those given by a bilingual speaker under the same constraint of encapsulating the specified words.

Moreover, the assessors' evaluations of the meaning preservation did not show significant differences among the three conditions according to the Kruskal–Wallis test ($p = 0.124$), as illustrated in Figure 8. The two one-sided tests with the l-correction [48] further supported the equivalence between any two pairs of the three conditions. Thus, it is suggested that replacing the original phrases with their translations produced by the proposed method would not bother users by presenting phrases losing semantic information.

From these points, we conclude that H1 is partially supported. Specifically, VocabEncounter can present the word usages with a human-comparable quality regarding their naturalness and meaning preservation. On the other hand, the naturalness of the usages could not be the best effort due to the constraint of containing the specified words. Still, we cannot deny the possibility that VocabEncounter can support vocabulary learning via presenting such usages. We thus decided to run online experiments to evaluate the effectiveness of VocabEncounter as an entire vocabulary-learning system, including the effect of presenting those usages during Web browsing, in the following sections.

6.5 Deciding the threshold for presenting phrases

To run VocabEncounter as an entire system, we need to decide the threshold th_2, which is used for suppressing translated phrases that do not hold naturalness or their original meaning (Section 3.3). For this purpose, we also examined the relationship of the score obtained by Sentence-BERT and translation models (Equation 1) to the assessors' evaluations. As presented in Figure 9, we found



*** The presentation order of phrases is shuffled for each assessor.**

Figure 6: Assignment of the three conditions across the six groups of the assessors. The presentation order of the phrases is also randomly shuffled for each assessor.

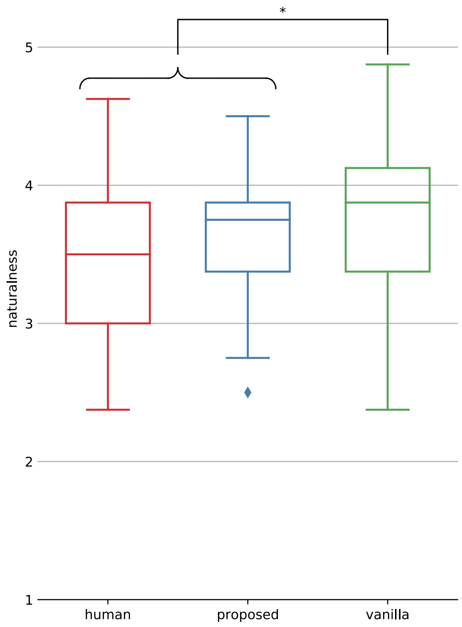


Figure 7: Comparison of the assessors' evaluations among the three conditions regarding the naturalness of the translated phrases. We found significant differences between the vanilla condition and the other conditions ($p < 0.05$).

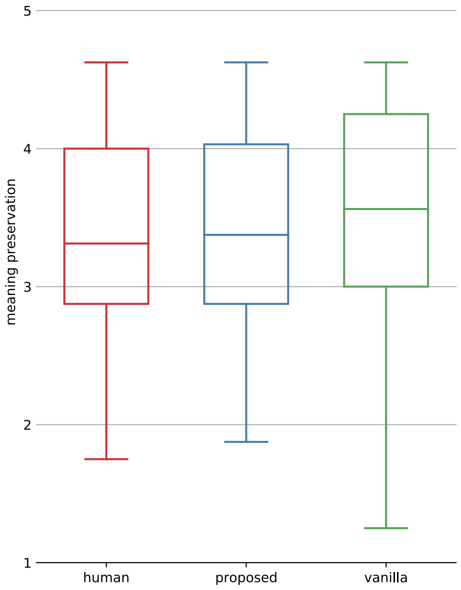


Figure 8: Comparison of the assessors' evaluations among the three conditions regarding the meaning preservation of the translated phrases. We found no significant difference among them.

that the score of each translated phrase correlated to the averaged evaluation of the phrase regarding its degree of meaning preservation ($\rho = 0.717, p < 0.001$) while its correlation to the naturalness of the phrase was not significant ($p = 0.077$). Considering that the Sentence-BERT model is trained to score the semantic similarity between given sentences and its output is weighted in Equation 1, this result is reasonable and consistent with our expectation.

Here, we have to be aware of the trade-off between the quality of the translated phrases to be presented and the chances for each word of W^{Ja} to be encapsulated in D^{Ja}, which is moderated by th_2. The higher threshold value leads to better translation results while allowing users to encounter fewer words they want to learn. In this case, we concluded that $th_2 = 0.6$ is reasonable. According to Figure 9, this value yields translations whose naturalness would be evaluated better than the neutral option of the Likert scale on average. In addition, using the data we collected in Section 4.1, we confirmed that the recall of the W^{En} was 0.752 when $th_2 = 0.6$. In other words, an average user is presumed to encounter approximately 75% of the words in an encapsulated manner when they use VocabEncounter for a day.

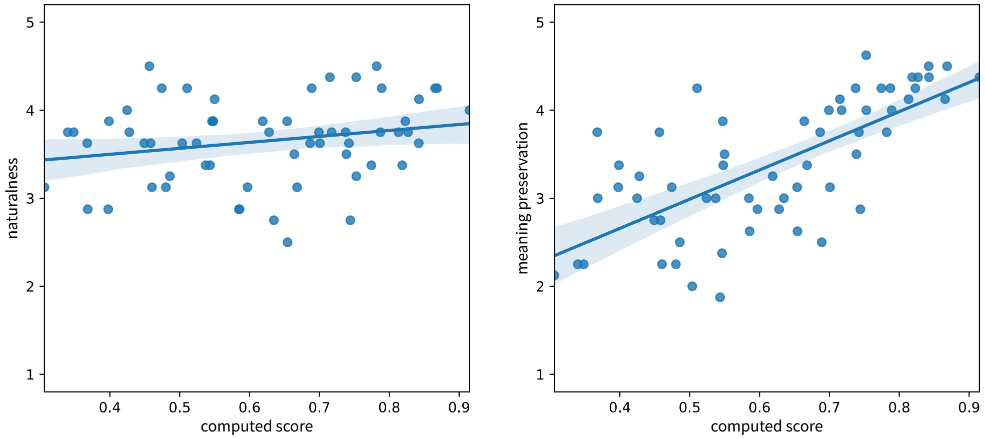


Figure 9: Relationships of the score obtained by Sentence-BERT and translation models (Equation 1) to the assessors' evaluation on the naturalness (left) and meaning preservation (right).

7 EXPERIMENT II: EFFICACY ON LEARNING OUTCOME

We next conducted a user study to examine whether VocabEncounter, as in the example scenario in Section 1, can help users memorize foreign words. To evaluate this point, we gathered participants and had them use VocabEncounter on their PCs for a day while browsing the Web as usual. Here, as we discussed in Section 5 to introduce H2, we expect that VocabEncounter helps users learn words, particularly by presenting their generated usages (i.e., a *phrase-based* interface). To verify this point, we prepared a *word-based* interface for comparison. It simply replaces the word in the original content, which is identified in the targeting mechanism (Section 3.2), with the word to remember (Figure 10). As we mentioned in Section 2.1, previous works have shown the efficacy of such a word-level translation on vocabulary learning, leveraging the strategy of micro learning. Therefore, H2 is supported if the *phrase-based* condition leads to a better learning outcome than the *word-based* condition.

7.1 Design

We used a within-participant design comparing the prepared conditions and examined their learning outcome. Here, to be a fair comparison, we needed to carefully design the words that each participant learn in this experiment. In particular, the words should be unknown to the participant and have the same difficulty (i.e., vocabulary level). Thus, we first asked each participant to take a pretest to identify words that they did not know from the same CEFR level, as detailed later in Section 7.3.

They then spent a day using the Chrome extension while each identified word was assigned to each experimental condition. Here, in addition to the *phrase-based* condition and the *word-based* condition, we added the *not-used* condition. This *not-used* condition was introduced as a baseline to measure the learning outcome without using the proposed system, i.e., how the participant could learn a word from a one-shot exposure at the pretest. More specifically, if a word was assigned to the *phrase-based* condition or the *word-based* condition, the participant could encounter them in a way shown in Figure 10, respectively. On the other hand, if a word was assigned to the *not-used* condition, the word was never shown to the participant. The words identified as unknown to the participant at the pretest were then randomly assigned to each of the three conditions, and this assignment did not change during the experiment. Note that, although we mentioned that users could add and delete words to remember when using VocabEncounter (Section 3.4), we deactivated this feature to keep the words in each condition the same during the experiment.

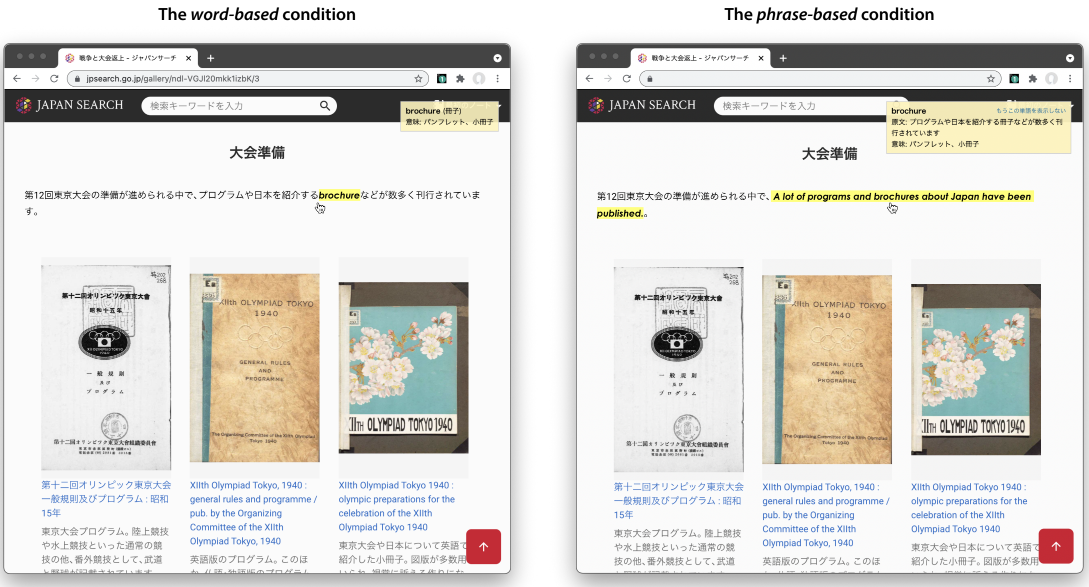
7.2 Measure

In order to measure the learning outcome of a participant, we used a vocabulary test and examined their correct answer rate on the words identified as unknown in the pretest. Specifically, in a similar manner as [90], we had them take a posttest two days after using the Chrome extension. By comparing how many of the words they correctly answered between the experimental conditions, we evaluated how VocabEncounter contributed to the memorization of foreign words.

7.3 Procedure

We gathered ten participants (three women and seven men, ages ranging from their 20s to 40s.) through word-of-mouth and online recruiting, who were native Japanese speakers and self-reported that their English level was equivalent to or slightly below CEFR B2. This level of mastery was selected to assure there would be some unknown words in the word list corresponding to B2, which motivated us to use the same list in Section 4.1.

The procedure of our experiment simulates the example scenario described in Section 1: the participants first tried to memorize unknown words; they spent a typical day using VocabEncounter; and their word memorization was tested. Figure 11 presents the actual procedure we used.


The *word-based* condition

戦争と大会返上 - ジャパンサーチ

jpsearch.go.jp/gallery/ndl-VGJl20mkk1izbK/3

JAPAN SEARCH

検索キーワードを入力

brochure (冊子)
意味: パンフレット、小冊子

大会準備

第12回東京大会の準備が進められる中で、プログラムや日本を紹介する*brochure*などが数多く刊行されています。

第十二回オリンピック東京大会一般規則及びプログラム : 昭和15年

東京大会プログラム。陸上競技や水上競技といった通常の競技の他、番外競技として、武道

XIIth Olympiad Tokyo, 1940 : general rules and programme / pub. by the Organizing Committee of the XIIth Olympiad Tokyo, 1940

英語版のプログラム。このほ

XIIth Olympiad Tokyo 1940 : olympic preparations for the celebration of the XIIth Olympiad Tokyo 1940

東京大会や日本について英語で紹介した小冊子。図版が多数用

The *phrase-based* condition

戦争と大会返上 - ジャパンサーチ

JAPAN SEARCH

検索キーワードを入力

brochure もうこの単語を表示しない
原文: プログラムや日本を紹介する冊子などが数多く刊行されています
意味: パンフレット、小冊子

大会準備

第12回東京大会の準備が進められる中で、*A lot of programs and brochures about Japan have been published.*。

第十二回オリンピック東京大会一般規則及びプログラム : 昭和15年

東京大会プログラム。陸上競技や水上競技といった通常の競技の他、番外競技として、武道

XIIth Olympiad Tokyo, 1940 : general rules and programme / pub. by the Organizing Committee of the XIIth Olympiad Tokyo, 1940

英語版のプログラム。このほ

XIIth Olympiad Tokyo 1940 : olympic preparations for the celebration of the XIIth Olympiad Tokyo 1940

東京大会や日本について英語で紹介した小冊子。図版が多数用

Figure 10: For comparison, we prepared a *word-based* interface (left) that simply replaces a word instead of presenting a translated phrase (right).

As we mentioned in Section 7.1, we first had the participants take a pretest to identify unknown words (Figure 11A). Here, words from the B2 word list were shown to each participant in a random order along with five options for its meaning written in Japanese (including one indicating "I do not know this word"). The participant was asked to choose one of the five options. The words were shown until the participant answered incorrectly or indicated lack of knowledge 60 times in total.

After 60 unknown words were identified, they were then randomly assigned to each of the three conditions so that each condition had 20 words. In addition, their correct answers were shown so that the participant could know their meanings. This was intended to perform a fair comparison within the three conditions; that is, the participant might never encounter some words assigned to the *phrase-based* and *word-based* conditions during the experiment (see Section 6.5). Thus, without this process, the participant would have no chance to know the meaning of such words, which makes the comparison of the correct answer rate across the two conditions and the *not-used* condition improper.

On the experiment day, we first had the participant watch a video describing how to use VocabEncounter. Then, they installed VocabEncounter in their PCs and spent the day as usual while encountering words during Web browsing (Figure 11B). Two days after the experiment day, we asked the participant to take a posttest (Figure 11C), which had the same format as the pretest but only presented the 60 words they had answered incorrectly in the pretest. The participants were given approximately $50 as a reward for their participation. Note that we conducted the experiment remotely for all participants.

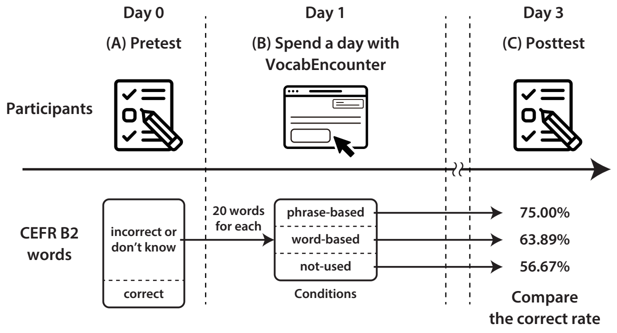


Figure 11: Procedure of the experiment. (A) Participants first took the pretest, and the words they did not know were used as W^{En}. (B) On the experiment day, they used VocabEncounter while they browsed the Web as usual. (C) Two days after the experiment day, they took the posttest to examine whether they had learned the words.

7.4 Results

Figure 12 shows the correct answer rate across the three experimental conditions. Our repeated-measures ANOVA analysis indicated a significant difference between the conditions ($F = 7.909, p = 0.003$), and the post-hoc test showed that the *phrase-based* condition resulted in a significantly better rate than the *word-based* condition ($p = 0.046$) and the *not-used* condition ($p = 0.001$). These results confirm that VocabEncounter helped the participants learn foreign words without requiring them to take dedicated time for studying. In particular, its presentation of the generated usages enhanced the

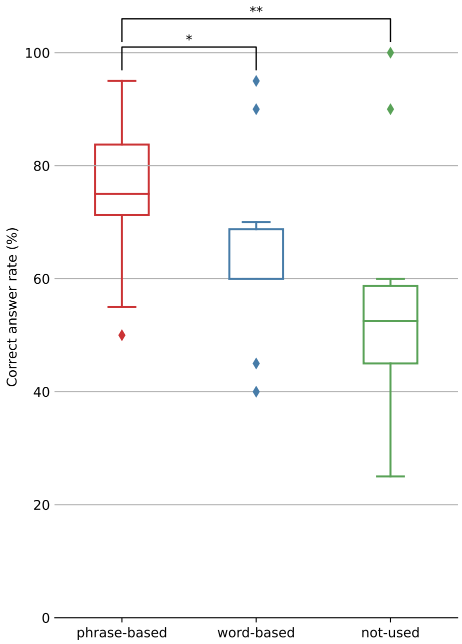


Figure 12: Comparison of the correct answer rate achieved by the participants. The *phrase-based* condition exhibited a significantly better rate than the *word-based* condition (p = 0.046) and the *not-used* condition (p = 0.001).

learning outcome, compared to presenting only the word. From these points, we conclude that H2 is supported.

8 EXPERIMENT III: LONG-TERM USER STUDY

So far, we have shown that VocabEncounter can generate natural usages of specified words without losing the original context (Section 6) and the generated usages can induce a better learning outcome than a word-level translation (Section 7). Finally, we conduct a user study for a more extended period simulating the envisioned scenario (Section 1). By doing so, we qualitatively evaluate the usability of VocabEncounter through semi-structured interviews and discuss its design implications in this section. Here, H3 is supported if participants favor VocabEncounter and are willing to use it in their daily lives.

8.1 Procedure

We gathered five participants (two women and three men, age ranging from their 20s to 30s) in the same manner as described in Section 7.3. None of them had participated in the second experiment (Section 7). After agreeing to the research policy, they watched an instruction video presenting the usage of VocabEncounter. Then, they installed VocabEncounter and spent one week while using it freely. Upon starting the one-week use, we asked them to register some English words they wanted to remember on VocabEncounter, such as ones they had read in a vocabulary book or made incorrect answers in a vocabulary test in recent but did not remember their meanings yet. Moreover, the addition and deletion function of VocabEncounter was activated, in contrast to Section 7, so that the participants could customize the words to learn on their own demand during the week.

After a participant spent a week, an experimenter conducted semi-structured interviews to understand how they found the use of VocabEncounter in terms of its usability and efficacy on learning outcomes. The participants were asked a series of questions, including “Could you tell us your overall impressions of using the system?”; “What do you think are the advantages of using the system?”; “What do you think are the disadvantages of using the system?”; “Did you have any words that you could remember or presentations that left an impression on you?”; “How do you think we can improve the system?”; and “Do you want to continue using the system?” Then, they were given approximately $50 as a reward for their participation. Note that we conducted the experiment remotely for all participants.

8.2 Results

By analyzing the participants' responses using open coding [74], we identified three major topics regarding the usability and efficacy of VocabEncounter. Overall, they favored the experience of using VocabEncounter, appreciating micro and usage-based learning enabled by the system, as we anticipated regarding H3. At the same time, they mentioned several points to improve VocabEncounter, which informed us how we could further leverage computers for vocabulary learning.

8.2.1 Benefits of micro learning. When we asked them for their impression about learning vocabulary with VocabEncounter, all participants mentioned the advantage of micro learning, giving concrete examples they encountered:

> I added the word *clarify* because I was often confused in remembering its spelling. Thanks to the repetitive presentation, I am now confident of the spell. (P4)

> When I was informed of the function of the system, I thought that it might be annoying. But, I never felt so; rather, it gave me opportunities to encounter English technical terms that I would not even recall without this system, such as *respiratory*. (P2)

These comments confirm the findings from the existing techniques enabling micro learning (Section 2.1) and rationalize our design of VocabEncounter delivering this learning strategy.

The participants provided an additional perspective regarding why they felt this learning strategy is effective:

> I sometimes ignored the presented translations and immediately referred to their original phrases, prioritizing reading contents on Web pages than learning foreign words. Still, seeing such translations during browsing reminded me that I have words to learn. (P2)

> I'm routinely studying new English words for the TOEFL test, but actually, I often feel reluctant to take

> time. In this respect, it was good to have opportunities to encounter and remember the words naturally without actively taking time for studying. (P5)

As discussed in Section 2.1, micro learning is known to be effective in allowing learners to leverage small learning units [30, 40]. These comments suggest that our approach of exploiting daily browsing further motivated them through exposure to foreign words in their daily lives.

8.2.2 Benefits of usage-based learning. The participants also appreciated that VocabEncounter enabled them to remember foreign words with their usages. In particular, they mentioned the benefit of being presented usages that were generated from the contents they were reading, as follows:

> It was nice that this system allowed me to learn phrases associated with topics I am interested in (for example, movies or fashion industry) through presenting usages when I visited Web sites about the topics. (P1)

> In comparison to reading translated phrases in existing learning materials, it was quite interesting to see how phrases on the page I was reading were translated. Such unpredictable translations on the page provided me with a clue to remember the word. (P3)

These advantages of presenting usages grounded in a user's context are, as discussed in Section 2.2, challenging to achieve without the NLP techniques we adopted.

Given these benefits, all participants responded that they want to keep using VocabEncounter, such as:

> I want to keep using it because it does not require any extra effort while it provides an efficient way to learn new foreign words. (P4)

These comments confirm that the participants favored the learning experience of VocabEncouter, which benefited from both micro and usage-based learning, after their one-week use of the proposed system. Therefore, we conclude that H3 is supported.

8.2.3 Room for providing further efficient vocabulary learning. At the same time, the participants mentioned several points to improve VocabEncounter, which informed us how we could construct better interaction techniques for vocabulary learning. Some comments provided ways to improve the presentation of VocabEncounter; for example, one participant suggested highlighting the encapsulated word when translated phrases are presented.

Another participant suggested implementing a feature to show their learning history:

> I would like to see a summary page where each word I tried to remember is listed with its usage I actually encountered. Then, I would be able to remember the words more efficiently by reflecting on them. (P2)

In fact, the literature on language learning confirms the effectiveness of the suggested feature, as Tseng and Schmitt [82] reported that reflecting one's learning process induced an improvement on their performance of vocabulary learning. In conventional vocabulary learning, learners can easily reflect by referring to the materials they used (e.g., vocabulary book). On the other hand, as the usages presented by VocabEncounter are dynamically generated during browsing, users would not be able to review their learning history without such a summarizing feature. This comment suggests that we should consider allowing users to reflect when we employ on-the-spot generation techniques for vocabulary learning.

In addition, several participants mentioned that some foreign words were presented more frequently than other words.

> While the words *drench* and *sanity* appeared many times, I did not encounter the word *hectare* very often although I registered it as a word to remember. (P3)

> I found that some of the words I wanted to remember that did not match the content of the pages I visited did not appear as often. (P4)

As commented, there would be a specific affinity between the words and the documents, which can induce the imbalanced presentation of the words depending on the pages the participant visited. Moreover, some English words with a more general meaning (e.g., *drench* and *sanity*) may exhibit higher similarities with more Japanese words in multilingual word embedding and be likely to be used for the encapsulation. We discuss how we can overcome this point later in Section 11.1.

9 EXAMPLE USE CASES

So far, we have shown that VocabEncounter can successfully support users in memorizing new words during their Web browsing experiences via a Chrome extension. However, the proposed encapsulation approach is not only applicable to a browser extension but also to other situations. In this section, we demonstrate various examples of the proposed approach to be used for vocabulary learning throughout daily activities. (see Figure 13).

First, our implementation of VocabEncounter built upon a Chrome extension allows users to use all their time for vocabulary learning as long as they are using the browser. For instance, as depicted in Figure 13A, users can learn vocabulary while communicating with their colleagues on Slack. Considering that Brooks [11] emphasized the importance of the workplace context for adult learners and the difficulty of preparing such contextualized training, this use case envisions the possibility that VocabEncounter promotes vocabulary learning of people in various situations.

In addition, VocabEncounter can be used with smartphones by developing an app with WebView. This allows users to, for example, learn vocabulary while reading news articles on their smartphones during a commute on public transportation (Figure 1B). In other words, this integration with smartphones would further encourage users to take advantage of small time segments for vocabulary learning, as previous work explained in Section 2.1.

Moreover, we can exploit the time spent watching movies using VocabEncounter, as illustrated in Figure 1C. This use case is inspired by existing interaction techniques utilizing the scenes and subtitles of movies (see Section 2.2). However, different from them, VocabEncounter does not require users to watch the movies in a foreign language or with foreign subtitles. Instead, it shows translated subtitles containing the words they want to remember only for a small portion of the movies while watching it in their native language.

[6] The news article used is distributed under the CC BY 2.0 by NHN Japan Corp.

[7] The movie used is distributed under the CC BY 3.2 by WebTV ASO.

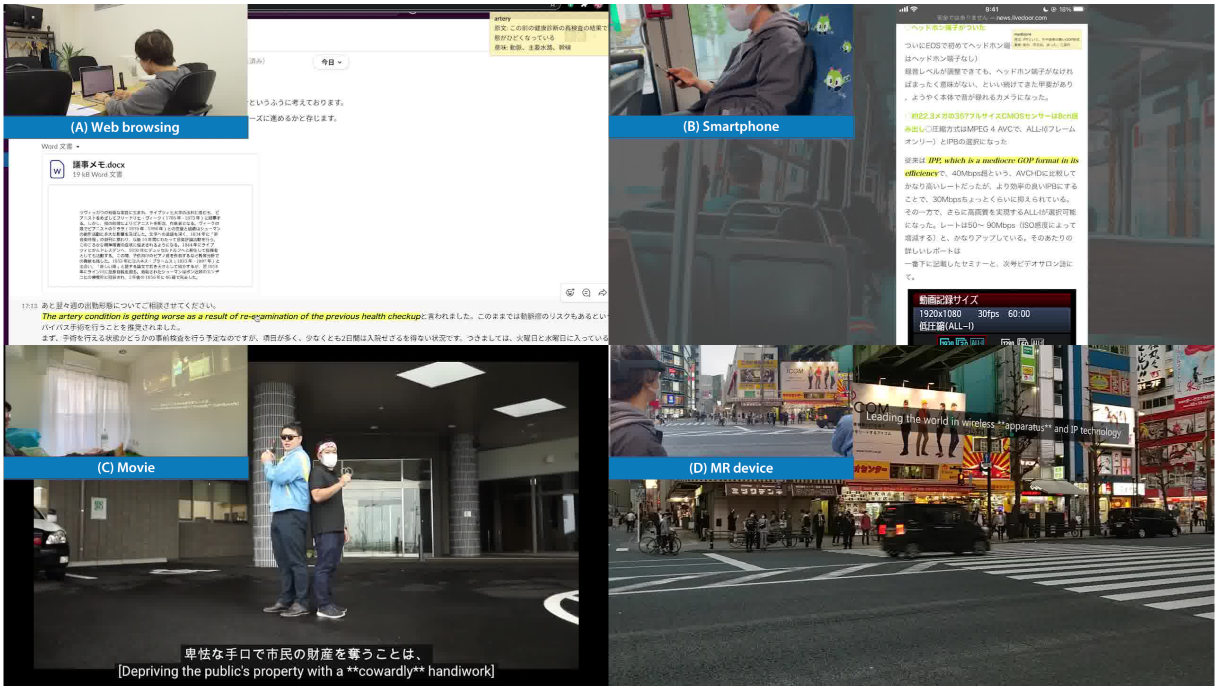

(A) Web browsing
(B) Smartphone
(C) Movie
(D) MR device
The artery condition is getting worse as a result of re-examination of the previous health checkupと言われました。
従来は IPP, which is a mediocre GOP format in its efficiencyで、40Mbps超という、AVCHDに比較して
卑怯な手口で市民の財産を奪うことは、
[Depriving the public's property with a **cowardly** handiwork]
Leading the world in wireless **apparatus** and IP technology

Figure 13: Example use cases of how VocabEncounter can transform our daily activities into a field of vocabulary learning. (A) The user can encounter the usages of the words they want to remember during Web browsing. (B) VocabEncounter also allows users to utilize the time they spend reading news on their smartphone during a commute[6]. (C) As it can be used while watching movies, the user can learn the words without dedicated study time[7]. (D) Mixed reality would further enable us to leverage our daily activities for vocabulary learning, such as strolling in the city.

Furthermore, it is possible to incorporate VocabEncounter with mixed reality, as exhibited in Figure 1D using Hololens. In this case, a cloud OCR API is used to detect texts from the camera, and the texts are used for encapsulating the words to learn. Then, the translated phrases produced by VocabEncounter to contain the words are overlaid on the original texts in the physical world using Hololens. In contrast to Vazquez et al. [85] that can present only pre-installed example usages, our approach enables users to leverage diverse text information around them for vocabulary learning.

These use cases show the extensive applicability of VocabEncounter in transforming users' daily lives into a field for vocabulary learning. We further discuss planned future work that can technically expand the bandwidth of our vision in Section 10.2.

10 FUTURE DIRECTIONS

Our results and example use cases demonstrated that our proposed approach of encapsulating the words a user wants to learn into their daily lives could be an effective way to learn vocabulary. At the same time, the results informed us how we could improve VocabEncounter to make it more usable and widely applicable. In this section, we discuss future directions of VocabEncounter regarding its technical aspects.

10.1 Making the threshold customizable

As discussed in Section 6.5, there is a trade-off between the quality (i.e., the naturalness and meaning preservation of translated phrases) and the recall (i.e., the chances for each foreign word to be encapsulated). Furthermore, this trade-off is controlled by the threshold th_2.

Here, we confirmed that the recall calculated using the collected data in Section 6 correlated with th_2, as illustrated in Figure 14. This result, along with Figure 9, suggests that we can provide intuitive control of the threshold based on its relation to naturalness, meaning preservation, and recall. By building a regressive model between the threshold and the three parameters, users can adjust how often the words appear, seeing how the quality of translated phrases would be affected, without being aware of the threshold behind it. Similarly, users can set the desired level of naturalness or meaning preservation and then observe the estimated recall.

By implementing this feature, the users can use VocabEncounter more flexibly to transform their daily lives into a field for vocabulary learning. For example, they can set the threshold higher while working by specifying the higher value of the desired quality level on the interface, resulting in a relatively small number of natural phrases preserving the original meaning. If they are browsing the Web to pass the time, they can fully exploit the time by adjusting the recall higher, allowing themselves to see a larger number of translated phrases that may not be perfectly syntactically natural

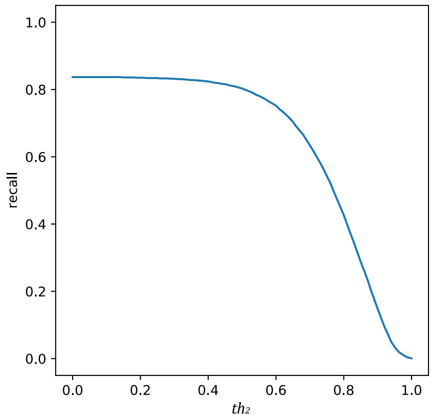


Figure 14: Relationship between th_2 and the recall calculated using the collected data in Section 6.

or exactly preserve the original meaning. As the importance of providing such a global control is emphasized regarding machine learning-based interactions [2], this customizing feature would contribute to the users' experience of VocabEncounter.

10.2 Expanding the application areas

In addition to the example use cases we demonstrated in Section 9, there are plans to improve VocabEncounter for expanding the vision of utilizing our everyday activities as a field of vocabulary learning. First, the proposed mechanism can be extended to encapsulate words of a specific language into texts written in the same language. For example, Satoru (the student featured in the example scenario in Section 1) can encounter the generated usages of English words he wants to learn while reading English materials. In this case, we first apply the targeting mechanism with d_i^{En} and W^{En} using monolingual word embedding of English instead of multilingual word embedding. Then, the targeted phrase is translated into $\tilde{p}_{i,j}^{Ja}$ with the help of an existing NMT algorithm. We apply the same constrained translation mechanism to $\tilde{p}_{i,j}^{Ja}$, obtaining $\tilde{p}_{i,j}^{En}$ that contains w_j^{En}. Then the original phrase in d_i^{En} is substituted with $\tilde{p}_{i,j}^{En}$ and presented to him. This extension of VocabEncounter makes the application scenario of watching movies (Figure 1C) more plausible, given that watching movies with captions in a foreign language is often considered an effective way to learn vocabulary [83]. In this case, some phrases in the captions are automatically substituted by VocabEncounter with ones containing the words to learn while maintaining the original meaning.

Secondly, the application scenario using the mixed reality we described in Section 9 can be enhanced in combination with neural caption-generation techniques [18, 87]. Using these techniques, we can automatically generate descriptions of scenes captured by Hololens while users spend a day wearing the device. VocabEncounter can leverage the generated descriptions for the encapsulation and present phrases describing the scenes in their daily lives while using a word they want to remember via the screen of Hololens. This extension would embody the two learning strategies we discussed in Section 2 (i.e., micro and usage-based learning) more faithfully, as it provides contextualized information more habitually by presenting usages of the words in a personalized manner.

11 LIMITATIONS

Although we have shown that VocabEncounter successfully helps users memorize new words through the proposed encapsulation approach, a few limitations remain with the current implementation. In this section, we discuss each point and possible future plans.

11.1 Word appearance

First of all, how often each word in W^{En} appears during Web browsing differs and relies on the contents users are reading, as we discussed in Section 8.2.3. We explained that this could be related to the characteristics of both the words to be encapsulated and the documents used to encapsulate the words. At the same time, in the context of machine learning, this could also be related to the *hubness* phenomenon [69], where some data points become k-nearest neighbors of many other points. This issue can be addressed by transforming the vector space of similarity [39, 57, 77] or simply suppressing the encapsulation of words if they appear too often within the same day. It is also possible to enable users to mark such words as remembered at any time so that VocabEncounter removes them from W^{En} in an interactive manner. The removal feature of VocabEncounter as we mentioned in Section 3.4 is also one such approach that allows users to control the appearance.

On the other hand, certain words may never appear during everyday browsing experiences when their meanings are uncommon, such as technical terms. VocabEncounter would be able to predict the likelihood of each word appearing during the daily browsing of average users by using the collected dataset of D^{Ja}, that is, examining whether each word would be used for the encapsulation at least once, as we did in Section 4.2. Thus, the system can recommend that users dedicate time to studying those words that have a low likelihood of appearing in their daily lives.

11.2 Syntactic validity

Another concern is that the NMT-based generation approach inevitably possesses a risk of presenting syntactically wrong phrases. In particular, while the naturalness of the translated phrases of the proposed method was confirmed to be human-comparable in Section 6.4, we found that the constraint of containing the specified word guided VocabEncounter to yield unnatural translations in some cases. Then, users might learn incorrect usages of the words while helping them memorize the meanings, especially when they are beginners.

As we discuss later in Section 11.6, more studies are expected to be conducted considering diverse situations in order to quantify the learning effects in depth. On the other hand, incorporating techniques for automatic grammatical error correction [31] would mitigate the risk. Such a post-check approach would also be effective to prevent the presentation of different words that is caused by the subword tokenizer, which we discussed in Section 6.4 by taking the example of "prom" and "promise."

11.3 User privacy

In addition, the users' privacy should be taken into account in the sense that a third-party server processes the content in their browser. In order to overcome this point, it is expected that VocabEncounter will be implemented in a privacy-preserving manner, such as implementing it as a client-side-only system with the help of TensorFlow.js [73] or adopting an NMT algorithm for secure multi-party computation [28].

11.4 Language pairs requiring consideration

Although we mentioned that VocabEncounter could be used for other pairs of languages in Section 3.1, for some languages, there may not be sufficient data to train the NLP models, i.e., models for multilingual word embedding and NMT. In addition, it should be considered that the languages of the translation pair can have different writing direction styles, such as English (left to right) and Arabic (right to left). In such a case, if a sentence in one language is mixed with a phrase in another language, it might interrupt the flow of users' reading. Thus, another form of presentation would be required.

11.5 Long-term learning outcome

Although the effectiveness of VocabEncouter in supporting memorizing new words was confirmed in Section 7, previous research [67] reported the existence of a forgetting effect in vocabulary learning. In this sense, it would be desirable to conduct a longer experiment to ensure its long-term learning outcome in comparison to conventional approaches. While Azadeh [62] reported that the difference in the short-term learning outcome caused by the alternation of learning strategies was widened in their long-term evaluation, such an extended experiment would further reveal the effect of transforming users' daily lives into a field for vocabulary learning.

11.6 In-depth learning outcome

More in-depth evaluations are demanded to explore the potential of VocabEncounter in supporting vocabulary learning throughout daily activities. For example, the vocabulary test we used for the pretest and posttest in Section 7 was insufficient to reflect the learning outcomes in depth. In particular, as our encapsulation approach enables users to encounter the practical usages of each word to remember, further studies are needed to quantify the learning effects in terms of how they can use the words practically.

In addition, our design of VocabEncounter implicitly assumes that users have a certain level of language skills to understand translated phrases. In this sense, it is desirable to expand the evaluation with users from different language levels to clarify the minimum level required to make VocabEncounter effective. Yet, considering that VocabEncounter translates phrases of approximately 10 words (see Section 3.2), the translated phrases would not be likely to involve complex grammatical structures, as we can infer from Table 1 (from Appendix). Thus, we believe that the required language skills would be eased.

Furthermore, exploring the possibility of incorporating VocabEncounter with other conventional learning strategies would be fruitful. For example, we presented a scenario of using a vocabulary book as a source of foreign words to learn in Section 1. In Section 7, we designed a procedure of having a vocabulary test to identify the words to learn with VocabEncounter. In addition, the design of VocabEncounter of transforming users' daily lives into a field for vocabulary learning does not prohibit its combined use with other existing interaction techniques we mentioned in Section 2.1 and Section 2.2. Therefore, examining the most effective way of incorporating several learning strategies would facilitate the application of VocabEncounter.

12 CONCLUSION

We proposed VocabEncouner, a novel vocabulary-learning system built on top of recent NLP techniques. We identified two key strategies for successful vocabulary learning through a literature review: *micro* and *usage-based* learning. To achieve both strategies simultaneously, VocabEncounter *encapsulates* foreign words that a learner wants to remember into their daily lives by presenting computer-generated usages of the words. Our approach leverages various techniques, such as multilingual word embedding and NMT with constrained decoding, to generate phrases containing the specified words. We first examined the quality of the generated phrases with crowdworkers and found that it is a human-comparable quality regarding their naturalness and meaning preservation. Then, we conducted a series of user studies where participants spent their lives with VocabEncounter in the form of a Chrome extension. The results showed that VocabEncounter helped them memorize words and its experiences are preferable, thanks to its design of achieving both micro and usage-based learning. Furthermore, we presented example use cases demonstrating how VocabEncounter can transform users' daily lives into a field of vocabulary learning. We believe that our implementations and results show how the generation techniques powered by machine learning can be exploited to achieve the idea of computer-mediated reality [51], enabling users to leverage their world for highly intellectual activity such as learning with the help of computers.

REFERENCES

[1] Mofareh Alqahtani. 2015. The Importance of Vocabulary in Language Learning and How to Be Taught. *International Journal of Teaching and Education* 3, 3 (2015), 21–34. https://doi.org/10.20472/te.2015.3.3.002

[2] Saleema Amershi, Daniel S. Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi T. Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 ACM CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, Article 3, 19 pages. https://doi.org/10.1145/3290605.3300233

[3] Peter Anderson, Basura Fernando, Mark Johnson, and Stephen Gould. 2017. Guided Open Vocabulary Image Captioning with Constrained Beam Search. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. ACL, Stroudsburg, PA, 936–945. https://doi.org/10.18653/v1/d17-1098

[4] Richard C. Anderson and Peter Freebody. 1981. Vocabulary knowledge. In *Comprehension and teaching: Research reviews*, John T. Guthrie (Ed.). International Reading Association, Newark, DE, 77–117.

[5] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In *Proceedings of the 3rd International Conference on Learning Representations*. ICLR, La Jolla, CA, 15 pages.

[6] Zhang Baicheng. 2009. Do Example Sentences Work in Direct Vocabulary Learning? *Issues in Educational Research* 19, 2 (2009), 175–189.

[7] Loïc Barrault, Magdalena Biesialska, Ondrej Bojar, Marta R. Costa-jussà, Christian Federmann, Yvette Graham, Roman Grundkiewicz, Barry Haddow, Matthias Huck, Eric Joanis, Tom Kocmi, Philipp Koehn, Chi-kiu Lo, Nikola Ljubesic, Christof Monz, Makoto Morishita, Masaaki Nagata, Toshiaki Nakazawa, Santanu Pal, Matt Post, and Marcos Zampieri. 2020. Findings of the 2020 Conference on Machine Translation (WMT20). In *Proceedings of the 5th Conference on Machine Translation*. ACL, Stroudsburg, PA, 1–55.

[8] Jennifer Beaudin, Stephen S. Intille, Emmanuel Munguia Tapia, Randy Rockinson, and Margaret E. Morris. 2007. Context-Sensitive Microlearning of Foreign Language Vocabulary on a Mobile Device. In *Proceedings of the 2007 European Conference on Ambient Intelligence*. Springer, Berlin, Germany, 55–72. https://doi.org/10.1007/978-3-540-76652-0_4

[9] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. 2003. A Neural Probabilistic Language Model. *Journal of Machine Learning Research* 3 (2003), 1137–1155.

[10] Daniel Berleant, Lingyun Shi, Xinxin Wei, Karthikeyan Viswanathan, Chinlin Chai, Nihad Majid, Yujiang Qu, and Prasad Sunkara. 1997. LEARN: Software for Foreign Language Vocabulary Acquisition from English Unrestricted Text. *Computer Assisted Language Learning* 10, 2 (1997), 107–120. https://doi.org/10.1080/0958822970100202

[11] Ann K. Brooks. 2009. Complexity and Community: Finding What Works in Workplace ESL. *New Directions for Adult and Continuing Education* 2009, 121 (2009), 65–74. https://doi.org/10.1002/ace.326

[12] Kimberly Anne Brooks-Lewis. 2009. Adult Learners' Perceptions of the Incorporation of Their L1 in Foreign Language Teaching and Learning. *Applied Linguistics* 30, 2 (2009), 216–235. https://doi.org/10.1093/applin/amn051

[13] Thomas S. Brown and Fred L. Perry. 1991. A Comparison of Three Learning Strategies for ESL Vocabulary Acquisition. *TESOL Quarterly* 25, 4 (1991), 655–670.

[14] Carrie J. Cai, Philip J. Guo, James R. Glass, and Robert C. Miller. 2015. Wait-Learning: Leveraging Wait Time for Second Language Education. In *Proceedings of the 2015 ACM CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 3701–3710. https://doi.org/10.1145/2702123.2702267

[15] Carrie J. Cai, Anji Ren, and Robert C. Miller. 2017. WaitSuite: Productive Use of Diverse Waiting Moments. *ACM Transactions on Computer–Human Interaction* 24, 1 (2017), 7:1–7:41. https://doi.org/10.1145/3044534

[16] José Camacho-Collados and Mohammad Taher Pilehvar. 2018. From Word To Sense Embeddings: A Survey on Vector Representations of Meaning. *Journal of Artificial Intelligence Research* 63 (2018), 743–788. https://doi.org/10.1613/jair.1.11259

[17] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Brian Strope, and Ray Kurzweil. 2018. Universal Sentence Encoder for English. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. ACL, Stroudsburg, PA, 169–174. https://doi.org/10.18653/v1/d18-2029

[18] Shaoxiang Chen, Ting Yao, and Yu-Gang Jiang. 2019. Deep Learning for Video Captioning: A Review. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. IJCAI Organization, San Mateo, CA, 6283–6290. https://doi.org/10.24963/ijcai.2019/877

[19] Ronan Collobert and Jason Weston. 2008. A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning. In *Proceedings of the 25th International Conference on Machine Learning*. ACM, New York, NY, 160–167. https://doi.org/10.1145/1390156.1390177

[20] David Dearman and Khai N. Truong. 2012. Evaluating the Implicit Acquisition of Second Language Vocabulary using a Live Wallpaper. In *Proceedings of the 2012 ACM CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 1391–1400. https://doi.org/10.1145/2207676.2208598

[21] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. ACL, Stroudsburg, PA, 4171–4186. https://doi.org/10.18653/v1/n19-1423

[22] Tilman Dingler, Corinna Giebler, Ulf Kunze, Tim Würtele, Niels Henze, and Albrecht Schmidt. 2016. Memory Displays: Investigating the Effects of Learning in the Periphery. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays*. ACM, New York, NY, 118–123. https://doi.org/10.1145/2914920.2915030

[23] Tilman Dingler, Dominik Weber, Martin Pielot, Jennifer Cooper, Chung-Cheng Chang, and Niels Henze. 2017. Language Learning On-the-Go: Opportune Moments and Design of Mobile Microlearning Sessions. In *Proceedings of the 19th ACM International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, New York, NY, 28:1–28:12. https://doi.org/10.1145/3098279.3098565

[24] Georgiana Dinu, Prashant Mathur, Marcello Federico, and Yaser Al-Onaizan. 2019. Training Neural Machine Translation to Apply Terminology Constraints. In *Proceedings of the 57th Conference of the Association for Computational Linguistics*. ACL, Stroudsburg, PA, 3063–3068. https://doi.org/10.18653/v1/p19-1294

[25] Darren Edge, Stephen Fitchett, Michael Whitney, and James A. Landay. 2012. MemReflex: Adaptive Flashcards for Mobile Microlearning. In *Proceedings of the 14th ACM International Conference on Human-Computer Interaction with Mobile Devices and Services – Demonstrations*. ACM, New York, NY, 193–194. https://doi.org/10.1145/2371664.2371707

[26] Darren Edge, Elly Searle, Kevin Chiu, Jing Zhao, and James A. Landay. 2011. MicroMandarin: Mobile Language Learning in Context. In *Proceedings of the 2011 ACM CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 3169–3178. https://doi.org/10.1145/1978942.1979413

[27] Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2020. Language-agnostic BERT Sentence Embedding. *arXiv* abs/2007.01852 (2020), 1–13 pages.

[28] Qi Feng, Debiao He, Zhe Liu, Huaqun Wang, and Kim-Kwang Raymond Choo. 2020. SecureNLP: A System for Multi-Party Privacy-Preserving Natural Language Processing. *IEEE Transactions on Information Forensics and Security* 15 (2020), 3709–3721. https://doi.org/10.1109/tifs.2020.2997134

[29] EF Education First. 2020. EF English Proficiency Index. http://www.ef.com/epi.

[30] Gerhard Gassler, Theo Hug, and Christian Glahn. 2004. Integrated Micro Learning–An Outline of the Basic Method and First Results. *Interactive Computer Aided Learning* 4 (2004), 1–7.

[31] Tao Ge, Furu Wei, and Ming Zhou. 2018. Reaching Human-Level Performance in Automatic Grammatical Error Correction: An Empirical Study. *arXiv* abs/1807.01270 (2018), 15 pages.

[32] Robert Godwin-Jones. 2011. Emerging Technologies: Mobile Apps for Language Learning. *Language Learning and Technology* 15, 2 (2011), 2–11.

[33] Peter Yongqi Gu. 2003. Vocabulary Learning in a Second Language: Person, Task, Context and Strategies. *The Electronic Journal for English as a Second Language* 7, 2 (2003), 1–25.

[34] Ralph Haefner. 1932. Casual Learning of Word Meanings. *The Journal of Educational Research* 25, 4–5 (1932), 267–277. https://doi.org/10.1080/00220671.1932.10880263

[35] Hany Hassan, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark, Christian Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, William Lewis, Mu Li, Shujie Liu, Tie-Yan Liu, Renqian Luo, Arul Menezes, Tao Qin, Frank Seide, Xu Tan, Fei Tian, Lijun Wu, Shuangzhi Wu, Yingce Xia, Dongdong Zhang, Zhirui Zhang, and Ming Zhou. 2018. Achieving Human Parity on Automatic Chinese to English News Translation. *arXiv* abs/1803.05567 (2018), 25 pages.

[36] Ari Hautasaari, Takeo Hamada, Kuntaro Ishiyama, and Shogo Fukushima. 2019. VocaBura: A Method for Supporting Second Language Vocabulary Learning While Walking. *Proceedings of the ACM on Interactive, Mobile, Wearable, and Ubiquitous Technologies* 3, 4 (2019), 135:1–135:23. https://doi.org/10.1145/3369824

[37] Chris Hokamp and Qun Liu. 2017. Lexically Constrained Decoding for Sequence Generation Using Grid Beam Search. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. ACL, Stroudsburg, PA, 1535–1546. https://doi.org/10.18653/v1/P17-1141

[38] J. Edward Hu, Huda Khayrallah, Ryan Culkin, Patrick Xia, Tongfei Chen, Matt Post, and Benjamin Van Durme. 2019. Improved Lexically Constrained Decoding for Translation and Monolingual Rewriting. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. ACL, Stroudsburg, PA, 839–850. https://doi.org/10.18653/v1/n19-1090

[39] Jiaji Huang, Qiang Qiu, and Kenneth Church. 2019. Hubless Nearest Neighbor Search for Bilingual Lexicon Induction. In *Proceedings of the 57th Conference of the Association for Computational Linguistics*. ACL, Stroudsburg, PA, 4072–4080. https://doi.org/10.18653/v1/p19-1399

[40] Theo Hug. 2005. Micro Learning and Narration: Exploring Possibilities of Utilization of Narrations and Storytelling for the Designing of "Micro Units" and Didactical Micro-Learning Arrangements. In *Proceedings of the 4th Media in Transition Conference*. Massachusetts Institute of Technology, Cambridge, MA, 13 pages.

[41] Michiel Joosse, Manja Lohse, Niels Van Berkel, Aziez Sardar, and Vanessa Evers. 2021. Making Appearances: How Robots Should Approach People. *ACM Transactions on Human-Robot Interaction* 10, 1, Article 7 (2021), 24 pages. https://doi.org/10.1145/3385121

[42] Nal Kalchbrenner and Phil Blunsom. 2013. Recurrent Continuous Translation Models. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. ACL, Stroudsburg, PA, 1700–1709.

[43] Ryan Kiros, Yukun Zhu, Ruslan Salakhutdinov, Richard S. Zemel, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. Skip-Thought Vectors. In *Proceedings of the 29th Annual Conference on Neural Information Processing Systems*. Curran Associates Inc, Red Hook, NY, 3294–3302.

[44] Cees M. Koolstra and Jonannes W. J. Beentjes. 1999. Children's Vocabulary Acquisition in a Foreign Language through Watching Subtitled Television Programs at Home. *Educational Technology Research and Development* 47, 1 (1999), 51–60. https://doi.org/10.1007/bf02299476

[45] Dejan Kovachev, Yiwei Cao, Ralf Klamma, and Matthias Jarke. 2011. Learn-as-you-go: New Ways of Cloud-Based Micro-learning for the Mobile Web. In *Proceedings of the 10th International Conference on Web-Based Learning*, Vol. 7048. Springer, Berlin, Germany, 51–61. https://doi.org/10.1007/978-3-642-25813-8_6

[46] Taku Kudo and John Richardson. [n.d.]. SentencePiece: A Simple and Language Independent Subword Tokenizer and Detokenizer for Neural Text Processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. ACL, Stroudsburg, PA, 66–71.

[47] Guillaume Lample, Alexis Conneau, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2018. Word Translation without Parallel Data. In *Proceedings of the 6th International Conference on Learning Representations*. ICLR, La Jolla, CA, 14 pages.

[48] Carolyn Lauzon and Brian Caffo. 2009. Easy Multiplicity Control in Equivalence Testing Using Two One-Sided Tests. *The American Statistician* 63, 2 (2009), 147–154. https://doi.org/10.1198/tast.2009.0029

[49] Mircea Filip Lungu, Luc van den Brand, Dan Chirtoaca, and Martin Avagyan. 2018. As We May Study: Towards the Web as a Personalized Language Textbook. In *Proceedings of the 2018 ACM CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 338. https://doi.org/10.1145/3173574.3173912

[50] Amanda Major and Tina Calandrino. 2018. Beyond Chunking: Micro-learning Secrets for Effective Online Design. *FDLA Journal* 3 (2018), 5 pages.

[51] Steve Mann. 2002. Mediated Reality with implementations for everyday life. Presence: Teleoperators and Virtual Environments – Online Companion. (Date Accessed: 30 March, 2021).

[52] Tomás Mikolov, Quoc V. Le, and Ilya Sutskever. 2013. Exploiting Similarities among Languages for Machine Translation. *arXiv* abs/1309.4168 (2013), 10 pages.

[53] Tomás Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Proceedings of the 27th Annual Conference on Neural Information Processing Systems*. Curran Associates Inc, Red Hook, NY, 3111–3119.

[54] Council for Cultural Co-operation Council of Europe Modern Languages Division, Education Committee. 2001. *Common European Framework of Reference for Languages: learning, teaching, assessment*. Cambridge University Press, Cambridge, UK.

[55] Jihyung Moon, Hyunchang Cho, and Eunjeong L. Park. 2020. Revisiting Round-trip Translation for Quality Estimation. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*. EAMT, Allschwil, Switzerland, 91–104.

[56] Makoto Morishita, Jun Suzuki, and Masaaki Nagata. 2020. JParaCrawl: A Large Scale Web-Based English-Japanese Parallel Corpus. In *Proceedings of The 12th Language Resources and Evaluation Conference*. ELRA, Paris, France, 3603–3609.

[57] Jiaqi Mu and Pramod Viswanath. 2018. All-but-the-Top: Simple and Effective Postprocessing for Word Representations. In *Proceedings of the 6th International Conference on Learning Representations*. ICLR, La Jolla, CA, 25 pages.

[58] William E. Nagy. 1995. *On the Role of Context in First-and Second-Language Vocabulary Learning*. Technical Report. Center for the Study of Reading, University of Illinois at Urbana-Champaign, Champaign, IL.

[59] Tatsuya Nakata. 2011. Computer-Assisted Second Language Vocabulary Learning in a Paired-Associate Paradigm: A Critical Investigation of Flashcard Software. *Computer Assisted Language Learning* 24, 1 (2011), 17–38. https://doi.org/10.1080/09588221.2010.520675

[60] Tatsuya Nakata. 2019. Learning Words With Flash Cards and Word Cards. In *The Routledge Handbook of Vocabulary Studies*. Routledge, Abingdon, UK, 304–320. https://doi.org/10.4324/9780429291586-20

[61] Paul Nation and Robert Waring. 1997. Vocabulary size, text coverage and word lists. In *Vocabulary: Description, Acquisition and Pedagogy*, Norbert Schmitt and Michael McCarthy (Eds.). Cambridge University Press, Cambridge, UK, 6–19.

[62] Azadeh Nemati. 2009. Memory Vocabulary Learning Strategies and Long-Term Retention. *International journal of vocational and technical education* 1, 2 (2009), 14–24.

[63] Nathan Ng, Kyra Yee, Alexei Baevski, Myle Ott, Michael Auli, and Sergey Edunov. 2019. Facebook FAIR's WMT19 News Translation Task Submission. In *Proceedings of the 4th Conference on Machine Translation*. ACL, Stroudsburg, PA, 314–319. https://doi.org/10.18653/v1/w19-5333

[64] Hiroaki Ogata, Chengjiu Yin, Moushir M. El-Bishouty, and Yoneo Yano. 2010. Computer Supported Ubiquitous Learning Environment for Vocabulary Learning. *International Journal of Learning Technology* 5, 1 (2010), 5–24. https://doi.org/10.1504/IJLT.2010.031613

[65] Mai Omura and Masayuki Asahara. 2018. UD-Japanese BCCWJ: Universal Dependencies Annotation for the Balanced Corpus of Contemporary Written Japanese. In *Proceedings of the 2nd Workshop on Universal Dependencies*. ACL, Stroudsburg, PA, 117–125. https://doi.org/10.18653/v1/w18-6014

[66] Myle Ott, Sergey Edunov, Alexei Baevski, Angela Fan, Sam Gross, Nathan Ng, David Grangier, and Michael Auli. 2019. fairseq: A Fast, Extensible Toolkit for Sequence Modeling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Demonstrations)*. ACL, Stroudsburg, PA, 48–53. https://doi.org/10.18653/v1/n19-4009

[67] Philip I. Pavlik and John R. Anderson. 2005. Practice and Forgetting Effects on Vocabulary Memory: An Activation-Based Model of the Spacing Effect. *Cognitive Science* 29, 4 (2005), 559–586. https://doi.org/10.1207/s15516709cog0000_14

[68] Matt Post and David Vilar. 2018. Fast Lexically Constrained Decoding with Dynamic Beam Allocation for Neural Machine Translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. ACL, Stroudsburg, PA, 1314–1324. https://doi.org/10.18653/v1/n18-1119

[69] Milos Radovanovic, Alexandros Nanopoulos, and Mirjana Ivanovic. 2010. Hubs in Space: Popular Nearest Neighbors in High-Dimensional Data. *Journal of Machine Learning Research* 11 (2010), 2487–2531.

[70] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. ACL, Stroudsburg, PA, 3980–3990. https://doi.org/10.18653/v1/D19-1410

[71] Nils Reimers and Iryna Gurevych. 2020. Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. ACL, Stroudsburg, PA, 4512–4525. https://doi.org/10.18653/v1/2020.emnlp-main.365

[72] Sebastian Ruder, Ivan Vulic, and Anders Søgaard. 2019. A Survey of Cross-lingual Word Embedding Models. *Journal of Artificial Intelligence Research* 65 (2019), 569–631. https://doi.org/10.1613/jair.1.11640

[73] Daniel Smilkov, Nikhil Thorat, Yannick Assogba, Ann Yuan, Nick Kreeger, Ping Yu, Kangyi Zhang, Shanqing Cai, Eric Nielsen, David Soergel, Stan Bileschi, Michael Terry, Charles Nicholson, Sandeep N. Gupta, Sarah Sirajuddin, D. Sculley, Rajat Monga, Greg Corrado, Fernanda B. Viégas, and Martin Wattenberg. 2019. TensorFlow.js: Machine Learning For The Web and Beyond. In *Proceedings of the 2nd SysML Conference*. mlsys.org, Indio, CA, 309–321.

[74] Anselm L Strauss snf Juliet M Corbin. 1990. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Sage Publications, Newbury Park, CA.

[75] Kai Song, Yue Zhang, Heng Yu, Weihua Luo, Kun Wang, and Min Zhang. 2019. Code-Switching for Enhancing NMT with Pre-Specified Translation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. ACL, Stroudsburg, PA, 449–459. https://doi.org/10.18653/v1/n19-1044

[76] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to Sequence Learning with Neural Networks. In *Proceedings of the 27th Annual Conference on Neural Information Processing Systems*. Curran Associates Inc, Red Hook, NY, 3104–3112.

[77] Ikumi Suzuki, Kazuo Hara, Masashi Shimbo, Marco Saerens, and Kenji Fukumizu. 2013. Centering Similarity Measures to Reduce Hubs. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. ACL, Stroudsburg, PA, 613–623.

[78] Rohail Syed and Kevyn Collins-Thompson. 2017. Retrieval Algorithms Optimized for Human Learning. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 555–564. https://doi.org/10.1145/3077136.3080835

[79] Rohail Syed and Kevyn Collins-Thompson. 2018. Exploring Document Retrieval Features Associated with Improved Short- and Long-term Vocabulary Learning Outcomes. In *Proceedings of the 41st International ACM SIGIR Conference on Human Information Interaction and Retrieval*. ACM, New York, NY, 191–200. https://doi.org/10.1145/3176349.3176397

[80] Yukio Tono and Masashi Negishi. 2012. The CEFR-J: Adapting the CEFR for English language teaching in Japan. *Framework & Language Portfolio SIG Newsletter* 8 (2012), 5–12.

[81] Andrew Trusty and Khai N. Truong. 2011. Augmenting the Web for Second Language Vocabulary Learning. In *Proceedings of the 2011 ACM CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 3179–3188. https://doi.org/10.1145/1978942.1979414

[82] Wen-Ta Tseng and Norbert Schmitt. 2008. Toward a Model of Motivated Vocabulary Learning: A Structural Equation Modeling Approach. *Language Learning* 58, 2 (2008), 357–400. https://doi.org/10.1111/j.1467-9922.2008.00444.x

[83] Robert Vanderplank. 2016. *Captioned Media in Foreign Language Learning and Teaching: Subtitles for the Deaf and Hard-of-Hearing as Tools for Language Learning*. Palgrave Macmillan, London, UK.

[84] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. In *Proceedings of the 31st Annual Conference on Neural Information Processing Systems*. Curran Associates Inc, Red Hook, NY, 5998–6008.

[85] Christian David Vazquez, Afika Ayanda Nyati, Alexander Luh, Megan Fu, Takako Aikawa, and Pattie Maes. 2017. Serendipitous Language Learning in Mixed Reality. In *Proceedings of the 2017 ACM CHI Conference on Human Factors in Computing Systems – Extended Abstracts*. ACM, New York, NY, 2172–2179. https://doi.org/10.1145/3027063.3053098

[86] Stefan Wellek. 2010. *Testing Statistical Hypotheses of Equivalence and Noninferiority* (2 ed.). Chapman and Hall/CRC, Boca Raton, FL. https://doi.org/10.1201/EBK1439808184

[87] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron C. Courville, Ruslan Salakhutdinov, Richard S. Zemel, and Yoshua Bengio. 2015. Show, Attend and

Tell: Neural Image Caption Generation with Visual Attention. In *Proceedings of the 32nd International Conference on Machine Learning*. JMLR, Cambridge, MA, 2048–2057.

[88] Florence W. M. Yip and Alvin C. M. Kwan. 2006. Online Vocabulary Games as a Tool for Teaching and Learning English Vocabulary. *Educational Media International* 43, 3 (2006), 233–249. https://doi.org/10.1080/09523980600641445

[89] Daniel Zeman, Jan Hajic, Martin Popel, Martin Potthast, Milan Straka, Filip Ginter, Joakim Nivre, and Slav Petrov. 2018. CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*. ACL, Stroudsburg, PA, 1–21. https://doi.org/10.18653/v1/k18-2001

[90] Yeshuang Zhu, Yuntao Wang, Chun Yu, Shaoyun Shi, Yankai Zhang, Shuang He, Peijun Zhao, Xiaojuan Ma, and Yuanchun Shi. 2017. ViVo: Video-Augmented Dictionary for Vocabulary Learning. In *Proceedings of the 2017 ACM CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 5568–5579. https://doi.org/10.1145/3025453.3025779

A EXAMPLE PHRASES PRODUCED BY VOCABENCOUTER IN COMPARISON WITH HUMAN TRANSLATIONS

In Table 1, we present some of the phrases obtained in Section 6. The English phrases in the *proposed* column were translated from the original Japanese phrases by VocabEncounter under the constraint of containing the corresponding English words. The phrases in the *human* column were translated by a bilingual translator under the same constraint, and those in the *vanilla* column were translated by the same NMT model to the *proposed* column without the constraint. As you can see in rows #1 to #3, some phrases in the *vanilla* condition can contain the specified word, as our targeting mechanism tried to find similar words in Japanese. At the same time, the other rows exhibit that the proposed translation mechanism contributed to presenting the usages of the specified word.

[8] Japanese often omits the subject of a sentence, which leads to the ambiguity in English translation.

Table 1: Some examples of the phrases we obtained in Section 6.

	Original phrase	Word to contain	Proposed	Human	Vanilla
#1	ウィキメディア・コモンズには、アリストテレスに関連するメディアおよびカテゴリがあります	media	Wikimedia Commons has media and categories related to Aristotle.	There are categories and media about Aristotle on Wikipedia Commons.	Wikimedia Commons has media and categories related to Aristotle.
#2	有料プランへのアップグレードの際、まずAsanaにログインした状態で	upgrade	When you upgrade to a paid plan, log in to Asana first.	When you want to upgrade to a paid plan, you first need to log on to Asana.	When upgrading to a paid plan, log in to Asana first.
#3	1988年にはCBSレコードを買収した	acquisition	The acquisition of CBS Records in 1988	The acquisition of CBS records in 1988	In 1988, he acquired CBS Records
#4	BLEビーコンを用いた屋内位置推定システム	estimated	Indoor location estimated system using BLE beacon	The indoor location estimated system with BLE beacon	Indoor Positioning System Using BLE Beacon
#5	専攻主任の中村先生の代理でお送りしています	superintendent	On behalf of Prof. Nakamura, the superintendent of the department	I am sending on behalf of Mr. Nakamura who is superintendent	On behalf of Prof. Nakamura, the director of the department
#6	もっと前向きな内容になるはずだったの	optimistic	It was supposed to be a more optimistic content.	It should be optimistic.	It was supposed to be a more positive content.
#7	自分にピッタリの商品を探し出すことができ	retrieve	You can retrieve the perfect item for yourself.	I can retrieve the item which is suitable for me.	You can find the perfect item for yourself.
#8	引っ張ることで広背筋、三角筋を鍛えることができる	pluck	By plucking, you can train the wide back and triangular muscles.	You can train your latissimus dorsi and deltoid muscles by plucking.	Pulling allows you to train the broad spine and triangular muscles.
#9	私と彼がデートを重ねられた	prom	I and he have been dating for a long time, and I'm promise.	I and he were going to prom sometimes.	I and he were dating.