Helping Mobile Learners Know Unknown Words through their Reading Behavior

Riku Higashimura
Osaka Prefecture University
Sakai, Osaka, Japan
higashimura@m.cs.osakafu-u.ac.jp

Andrew Vargo
Osaka Prefecture University
Sakai, Osaka, Japan
awv@m.cs.osakafu-u.ac.jp

Motoi Iwata
Osaka Prefecture University
Sakai, Osaka, Japan
iwata@cs.osakafu-u.ac.jp

Koichi Kise
Osaka Prefecture University
Sakai, Osaka, Japan
kise@cs.osakafu-u.ac.jp

ABSTRACT

Vocabulary acquisition is a fundamental part of learning a new language. In order to acquire new vocabulary, words with meanings that are unknown to the learner must be added to the language learning process. When searching for material in the target language, it is useful to know how much of a document is made up of currently unknown words. One simple way to estimate the unknown words in a document is to use the frequency of occurrence, which indicates the difficulty of the word. However, this approach can lead to missed unknown words. In this study, we aim to improve the accuracy of unknown word estimation by using reading activity data obtained from smartphone sensors and taking into account the individual learner's English reading behavior. We apply a novel user interface which allows us to improve estimation through reading behavior, without the use of eye-trackers.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in ubiquitous and mobile computing*.

KEYWORDS

language learning, mobile learning, education

ACM Reference Format:

Riku Higashimura, Andrew Vargo, Motoi Iwata, and Koichi Kise. 2022. Helping Mobile Learners Know Unknown Words through their Reading Behavior. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '22 Extended Abstracts), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3491101.3519620

1 INTRODUCTION

The introduction of smartphones has allowed language learners to make their learning sessions mobile and more flexible within their daily schedules. While this is already a significant advantage over paper-based textbooks and more stationary computing devices, mobile language learning applications can also allow for adaptive and personalized learning regiments in order to make them more effective for users [16]. Personalization can include customization of context, adaptive practice routines, and confidence estimation. One particular area that needs development for mobile users is in the identification of unknown words for the learner.

Unknown words present an interesting and difficult problem for language learners [2]. If there are not enough unknown words, the learner cannot quickly build up their vocabulary sets quickly. However, too many unknown words in a document can make the document frustrating to read and can harm global comprehension. In addition, this makes it difficult to have unplanned learning sessions. If reading material is able to be curated by a teacher or professional service, it may be possible to have effective lesson plans. However, the cost that would be incurred by users is significant. Thus, the goal here is to create low-cost solutions which can be implemented in mobile environments and which require no content curation.

One way to quickly and easily estimate the likelihood of an individual knowing a word is to measure word frequency [9]. Typically, the less common a word is used within a language corpus, the more unlikely it is for any individual to know it. This scales with ability and exposure to the language, and is even applicable to native language speakers. However, this method is not sufficient for universal use, as it can lead to gaps in knowledge and is difficult to moderate for users who have domain-specific language proficiency. Many learners may have vast knowledge in a subject area, but wide gaps of missing vocabulary in other areas. For instance, a language learner who watches super hero movies frequently may have totally different knowledge base words compared to a learner who only reads computer science papers.

In this paper, we use individual users' reading behavior to improve the estimation of unknown words. We created a user interface (UI) which provides a window on each individual line in the text. This isolation enables us to identify words that are unknown to the reader. This allows us to implement an effective technique that can be employed on convenient mobile devices. We incorporate reading activity data obtained from smartphone sensors, such as reading time, 3-axis acceleration of the smartphone (2Hz), and the number of times the smartphone is reread. This data allows us to more accurately identify unknown words for 17 out of 19 participants in

CHI '22 Extended Abstracts, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9156-6/22/04…$15.00
https://doi.org/10.1145/3491101.3519620

our study as compared to user independent features such as word frequency.

While the context and focus of this paper is on university age students learning English, these results show the potential of this type of estimation not only for learners of new languages, but for many types of learning. For instance, the same premise put forth in this work may be applied to young learners or students of technical and specialized domains, where word frequency plays a vital role and mobile device usage is desirable.

2 BACKGROUND

The estimation of unknown words has been explored from both a document-based approach as well as from a reading behavior approach. In this section, we briefly discuss these approaches and their relationship with our work.

2.1 Quantitative Document Features on Sentence Difficulty

While frequency gives a rough measure of how likely a word is to be known, it can be improved upon. There have been multiple methods developed to automatically model and predict the difficulty of sentences based on quantitative document features [1, 4, 15]. Brown et al. [1] use characteristics of sentence structure, such as the percentage of functional words and the percentage of words with more than seven letters, as quantitative document features. Here, functional words are words that express grammatical relations such as prepositions and conjunctions. Senter et al. proposed the Automated Readability Index (ARI) as a measure of the difficulty of sentences, which is considered to be highly reliable [15]. The ARI is an index proposed for automatic computer processing and is calculated by the equation (1), where α and β are the average number of characters per word (average word length) and the average number of words per sentence (average sentence length), respectively, in the target multiple sentences.

$$ARI = (4.71 \times \alpha) + (0.5 \times \beta) - 21.43 \quad (1)$$

This metric allows us to obtain document features related to the difficulty of a sentence with higher accuracy than using only the frequency of word occurrence. Therefore, we use ARI as a user independent feature in addition to basic frequency.

2.2 Unknown Word Estimation with Eyetrackers

User behavior can tell us a lot about an individual's comprehension and understanding. One such type of user behavior is eyegaze, which can be measured with an eye-tracking device [5, 6, 10, 12]. Garain et al. [6] proposed a method for estimating reader specific difficult words in a document using eyegaze information while reading on a PC. They recorded the eyegaze information of five learners and verified the accuracy of the estimation of reader specific difficult words. The results showed that for all words the recall rate was 99.0% and the recall rate was 33.4%. When proper nouns were excluded from the training and evaluation, the precision was improved to 99.0% and the recall was 39.6%.

A limitation to eye-tracking is that it is best used on stationary devices. Eye-tracking using the front camera of a smartphone is not accurate enough [11]. It is also not feasible to attach a more capable eye-tracker to a smartphone. This means that other methods need to be developed in order to estimate unknown words on mobile devices.

2.3 Relationship Between English Reading Ability and Reading Time

Another way to estimate English reading ability is from observing reading time of documents [13]. We can see the correlation between reading score and reading time on Benesse's English Communication Skills Test [1], which was administered to 239 first- and second-year university students in the year 2000. As a result, Naganuma reported that "difficulty" and "topic," which are linguistic factors of document readability, affect reading speed. For difficulty, there are criteria such as "level of vocabulary," "complexity of sentence structure," and "length of text," and for the topic, there are categories such as "humanities" and "science." This indicates that reading time is an effective indicator of learners' English reading ability. Therefore, we use reading time as a user dependent feature.

3 PROPOSED METHOD

Our proposed method identifies whether a word is an unknown word or not by using user independent features, such as ARI and user reading behavior data obtained from smartphone sensor data. The combination allows for a more accurate identification system for most users.

3.1 Special Quantity Acquisition

We implemented a mobile application that displays a few words in a document line by line on a smartphone screen as shown in figure 1(b). This allows readers to focus on each line of text in isolation, and prevents the reader from straying from the text or simply skipping the difficult words. One line on the smartphone screen is defined as one window. We then acquire the document features and reading activity data for each window. This allows us to obtain the reading activity data in as few word units as possible, while not losing the readability of the document.

As shown in the Table 1, we identified up a total of 29 features: No.1−13 are 13 document features including the features introduced in Section 2.1, No.14−16 are 3 features related to the reading time, No.17−28 are 12 features related to the acceleration of the smartphone, and No.29 is 1 feature related to the number of rereadings.

3.2 Approach

We extracted the obtained features from one window unit to one word unit. In this case, stop words such as articles and proper nouns were excluded. From this, we add the frequency of occurrence of one of the extracted words to the features obtained for each window shown in Table 1 as a feature used for estimation. Table 2 shows the results of the feature extraction and the features used for estimation.

Sequential Backward Floating Stepwise Selection [14] was used to select features with high estimation accuracy. We then used

[1] https://www.benesse.co.jp/gtec/

Table 1: Features to be acquired for each window

No	Features
1-2	Number of words, number of characters
3-4	Average word length, average text length
5	Automated Readability Index
6-7	Number and percentage of function words
8-9	Number and percentage of words with 7 or more letters
10-13	Individual, minimum, maximum, and average values of frequency of occurrence
14	Reading time
15-16	Reading time per word, per letter
17-19	Minimum value of the smartphone's {x, y, z}axis acceleration
20-22	Maximum value of the smartphone's {x, y, z}axis acceleration
23-25	Average value of the smartphone's {x, y, z}axis acceleration
26-28	Distribution of the smartphone's {x, y, z}axis acceleration
29	Number of readbacks

Table 2: Features used for estimation

Count	Feature Descriptions
1	Word frequency
12	Document features obtained by 1 window unit
3	Reading time related to 1 window unit gained
12	Acceleration related to the smartphone obtained in 1 window unit
1	Number of readbacks obtained per window unit

Support Vector Machine (SVM) with an RBF kernel. The estimation is done in the user-dependent manner.

4 EXPERIMENT

An experiment was conducted with 19 participants (10 males and 9 females). The total duration of the experiment was 4 hours, of which 3 hours were spent on the task of reading the document. One hour was spent on post-hoc questionnaires. The participants were rewarded with a payment of 4,000 JPY (Approximately 35 USD). The age of the participants was an average of 22.2 years with a standard deviation of 1.47 years. All of the participants in the study were either undergraduate or graduate students in a university. Ethics approval was obtained prior to the beginning of the study.

We used the Newsela dataset [2] in our experiments. Newsela is a web service that provides English teaching materials, and learners can change the difficulty level of the documents based on their English ability.

[2]https://newsela.com/

4.1 Acquisition of Linguistic Features and Reading Activity Data

Each participant used their own smartphone for the experiment. Figure 1 shows the reading procedure. First, the participants were asked to select a document with a title they were interested in from the document selection screen shown in Figure 1(a). This was done because interest in content has been shown to be tied to comprehension and learning [8]. In order to try to ensure that participants faced a reasonable amount of unknown words, we asked them to choose a document that was estimated to have around 15 unknown words.

The selected document was then displayed on the reading screen shown in Figure 1(b). The participants were asked to read text one window at a time, and the reading activity data for each window was thus obtained. The participants were asked to read the document at their own pace, including rereading, and were instructed to "read the whole document while trying to understand its meaning".

After reading the document, the participants were asked to select words whose meanings they did not know as an unknown words on the screen shown in Figure 1(c). In addition, the participants were asked to give a summary of the document and its difficulty level (1: very easy – 5: very difficult) verbally to check whether they were reading the document.

These tasks were repeated, and data for nine documents were obtained. To prevent the participants from losing concentration, a 10-minute break was taken after every three documents were read. In total, participants read a range of 631 to 1,794 windows (Avg = 1,255.9 and SD = 370.6), a range of 3,119 to 7,724 words (Avg = 6,012.3 and SD = 1,704.4) and a range of 47 to 246 unknown words (Avg = 123.2 and SD = 44.1).

After the completion of the experiment, we administered the Big5 test [7] to examine the personality traits of the participants, the learning style survey (ILS) [3], and a questionnaire. The questionnaire sought to find out how the participants felt about the UI and to gain further insight into how they held their smartphones while using the program.

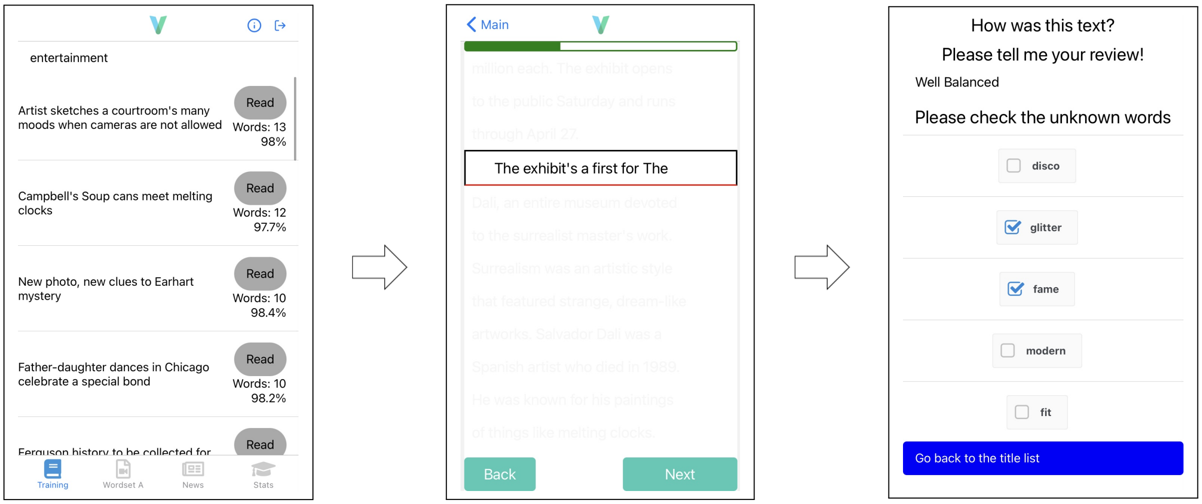
4.2 Evaluation Method

Cross-validation was performed in the user-dependent manner, and AUPR, the AUC of the PR curve, was calculated as an evaluation index. The AUPR was calculated based on the probability that a word was an unknown word, which was calculated when identifying whether a word was an unknown word or not for all the words in the nine documents for which features were extracted. We compared the AUPR of the proposed method with the baseline AUPR trained with only document features, which is an objective measure common to all learners.

To verify the accuracy of the estimation, we used Leave-One-Document-Out cross-validation on a document-by-document basis, using one of the documents as the test data and the other documents as the training data.

4.3 Results

The results of the comparison between the proposed method and the baseline AUPR are shown in Figure 2. This shows that the proposed method outperformed the baseline for 17 of the 19 experimental



(a) Document selection screen (b) Reading screen (c) Unknown word selection screen

Figure 1: Reading flow

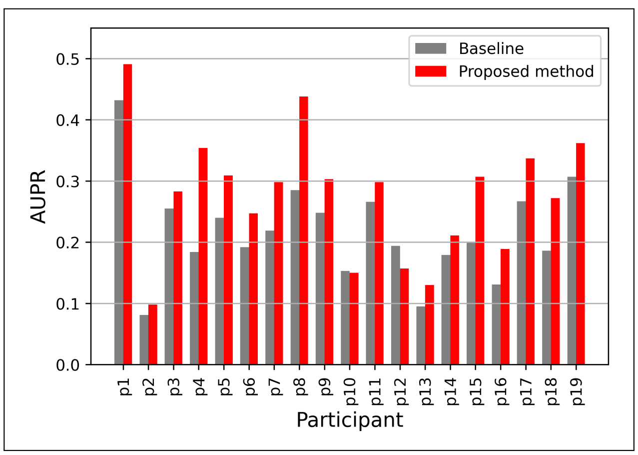


Figure 2: Proposed method and baseline AUPR

participants. A paired t-test (after confirmation that the data is normally distributed) was used to determine whether there was a significant difference between the baseline and the proposed method. The result, $p = 5.13\times10^{-5}$, shows that there is a statistically significant difference.

The average number of selected features for these 17 participants with positive results was 3.8. On the other hand, the number of selected features for the remaining two participants (p10 and p12) was 2 and 1, respectively, when the baseline outperformed the proposed method. This confirms that there were some participants whose reading activity data did not contribute to the estimation of the unknown words in this experiment. The post-hoc questionnaire indicated that these participants did felt that the words in the document were easy, indicating that the article that was selected was not of an appropriate difficulty level.

Next, we looked at the number of times a feature was selected as shown in Figure 3. The horizontal axis of Figure 3 shows the feature used for estimation and the vertical axis shows the number of times the feature was selected. The frequency of word occurrence was selected for all participants in the experiment. This confirms that word frequency is the feature that contributes the most to unknown word estimation, and provides a better baseline than ARI. The 3-axis acceleration (2Hz) of the smartphone also contributed to the unknown word estimation as much as the reading time, which indicates that how the smartphone is held can help with the estimation of unknown words.

5 DISCUSSION

In this paper, we provided a novel UI for estimating the number of unknown words in a document by using a combination of frequency and data acquired from smartphone usage. The overall results show that this method is effective at identifying unknown words for most users, surpassing just using word frequency. This gives an option to users to reduce the amount of manual work they need to do in order to identify unknown words. In addition, the UI is mobile and is able to be implemented without the use of an eye-tracking device.

There are some limitations. The method was not effective at all for two of the participants. In addition, the benefit was not particularly large for all of the users. As Figure 2 shows, some participants, such as p2 and p14, have modest gains over the baseline. This brings into question whether or not is worth the trouble and effort for such users to employ this system.

The results of Big5 and ILS questionnaires did not show any clear pattern that would explain why the method worked better for some participants compared to others. This may be because more participants are needed in order to see trends coalesce. An important step in future work is to better understand and predict for whom this type of technology will be effective, both in its current state and in the future. The post-hoc questionnaire did indicate that

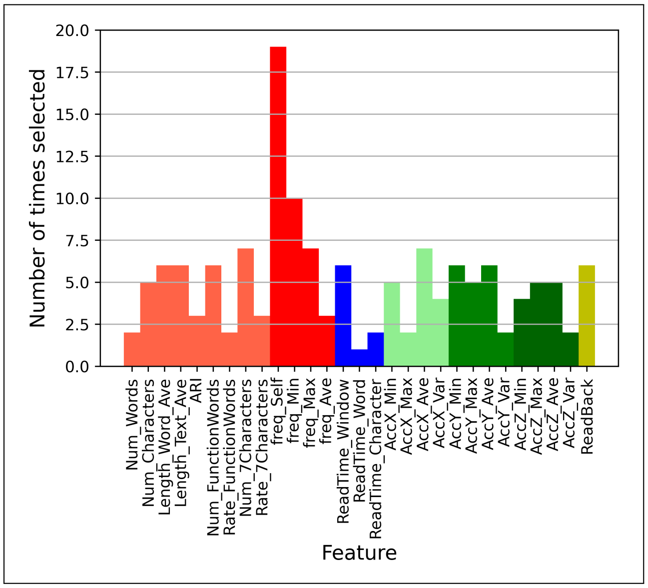


Figure 3: Number of times the feature used for estimation was selected

users did not feel strongly about the UI and were able to accept the layout.

We do believe that future work can improve this type of estimation. In our study, the sampling frequency of the 3-axis acceleration of the smartphone was 2Hz due to the memory processing of the smartphone. Therefore, if a mobile application with a higher sampling frequency is implemented, more detailed data on the 3-axis acceleration of the smartphone can be obtained, and it is likely better results can also be acquired.

There is also the potential to modify the UI in way which could be more universally beneficial to users. The current UI has a one-line interface that is uncomfortable for some participants to use. Therefore, the advantages of the system may be overcome by the UI. However, it maybe possible that a middle ground could be established with different interfaces that are more natural (e.g., highlighting line text). These could reduce some of the efficacy of the estimation, but may provide better user experiences that facilitate a worthy trade-off.

In addition, there is also the potential to extend this technique beyond foreign language learning, and into technical domains. This could help with the learning of jargon, or with confusion with words that have multiple meanings both within and outside of a domain. The identification of known words with unknown definitions in context will be an especially interesting challenge.

6 CONCLUSION

Conventional unknown word estimation methods use an objective measure common to all learners, such as the frequency of occurrence of a word, and thus cannot be optimized for individual users. User reading behavior can be used to personalize the estimation of unknown words. Because the state-of-the-art uses technology that is primarily available on stationary devices, there is a gap in accessibility for mobile users. To solve this problem, we developed a UI and proposed a method for estimating unknown words using reading activity data obtained from smartphone sensors.

ACKNOWLEDGMENTS

This research was supported in part by JST CREST (JPMJCR16E1), as well as JSPS Kakenhi (20H04213, 20KK0235).

REFERENCES

[1] James Dean Brown. 1997. An EFL readability index. *University of Hawai'i Working Papers in English as a Second Language* 15, 2 (1997), 85–119. http://hdl.handle.net/10125/40779

[2] James Coady and Thomas Huckin. 1997. *Second Language Vocabulary Acquisition: A Rationale for Pedagogy.* Cambridge University Press, Cambridge.

[3] Richard M Felder and Linda K Silverman. 1988. Learning and teaching styles in engineering education. *Engineering education* 78, 7 (1988), 674–681. https://doi.org/10.1016/0307-4412(88)90075-1

[4] Rudolph Flesch. 1948. A new readability yardstick. *Journal of applied psychology* 32, 3 (1948), 221. https://doi.org/10.1037/h0057532

[5] Katsuya Fujii and Jun Rekimoto. 2019. Subme: An interactive subtitle system with english skill estimation using eye tracking. In *Proceedings of the 10th Augmented Human International Conference 2019.* 1–9. https://doi.org/10.1145/3311823.3311865

[6] Utpal Garain, Onkar Pandit, Olivier Augereau, Ayano Okoso, and Koichi Kise. 2017. Identification of reader specific difficult words by analyzing eye gaze and document content. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Vol. 1. IEEE, 1346–1351. https://doi.org/10.1109/ICDAR.2017.221

[7] Lewis R Goldberg. 1992. The development of markers for the Big-Five factor structure. *Psychological assessment* 4, 1 (1992), 26. https://doi.org/10.1037/1040-3590.4.1.26

[8] Judith M. Harackiewicz, Jessi L. Smith, and Stacy J. Priniski. 2016. Interest Matters: The Importance of Promoting Interest in Education. *Policy Insights from the Behavioral and Brain Sciences* 3, 2, 220–227. https://doi.org/10.1177/2372732216655542

[9] Elfrieda H. Hiebert, Judith A. Scott, Ruben Castaneda, and Alexandra Spichtig. 2019. An Analysis of the Features of Words That Influence Vocabulary Difficulty. *Education Sciences* 9, 1, 8. https://doi.org/10.3390/educsci9010008

[10] Rui Hiraoka, Hiroki Tanaka, Sakriani Sakti, Graham Neubig, and Satoshi Nakamura. 2016. Personalized unknown word detection in non-native language reading using eye gaze. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction.* 66–70. https://doi.org/10.1145/2993148.2993167

[11] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. 2016. Eye tracking for everyone. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2176–2184. https://doi.org/10.1109/CVPR.2016.239

[12] Kai Kunze, Hitoshi Kawaichi, Kazuyo Yoshimura, and Koichi Kise. 2013. Towards inferring language expertise using eye tracking. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems.* 217–222. https://doi.org/10.1145/2468356.2468396

[13] Naganuma and Wada. 2002. Measurement of English Reading Ability by Reading Speed and Text Readability. *The Japan Language Testing Association* 5 (2002), 34–45. https://doi.org/10.20622/jltaj.5.0_34

[14] Pavel Pudil, Jana Novovičová, and Josef Kittler. 1994. Floating search methods in feature selection. *Pattern recognition letters* 15, 11 (1994), 1119–1125. https://doi.org/10.1016/0167-8655(94)90127-9

[15] RJ Senter and Edgar A Smith. 1967. *Automated readability index.* Technical Report. CINCINNATI UNIV OH. https://apps.dtic.mil/sti/citations/AD0667273

[16] Kohei Yamaguchi, Motoi Iwata, Andrew Vargo, and Koichi Kise. 2020. Mobile vocabulometer: a context-based learning mobile application to enhance English vocabulary acquisition. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers.* 156–159. https://doi.org/10.1145/3410530.3414406