Modelling second language learners for learning task recommendation

Haoran Xie
Department of Mathematics and Information Technology,
The Education University of Hong Kong, Hong Kong
Email: hxie@eduhk.hk

Di Zou
Department of English Language Education,
The Education University of Hong Kong, Hong Kong
Email: dzou@eduhk.hk

Tak-Lam Wong*
Department of Computing Studies and Information,
Douglas College, Canada
Email: wongi5@douglascollege.ca
*Corresponding author

Fu Lee Wang
Office of the President,
Caritas Institute of Higher Education, Hong Kong
Email: pwang@cihe.edu.hk

Abstract: How to recommend appropriate and effective learning tasks based on the characteristics of a second language learner is a vital question in the field of second language acquisition. In this research, we investigate the issue by dividing it into two sub-questions: how to model the characteristics of language learners as different learners may have varied expertise on and subjective preferences of many topics; and how to select learning tasks according to the constructed learner model. Research on the second sub-question has been widely conducted in domains such as recommender systems, and we focus on the first sub-question in this study from the perspective of how to model the preferred learning contexts of a learner in a non-intrusive manner. We conducted an experiment among eighty-two students, and the results showed that our proposed framework outperformed other systems as it provides significantly more effective and enjoyable word learning experience.

Keywords: learner modelling; context familiarity; task recommendation; word learning; e-learning.


Copyright © 2018 Inderscience Enterprises Ltd.
1 Introduction

How to recommend suitable learning tasks according to the characteristics of a second language learner is a vital question in the research community of the second language learning. A suitable language learning task can significantly facilitate the learning
effectiveness (Zou et al., 2014). This important research question can be further divided into two sub-questions. The first one is how to model the characteristics of second language learners, since different learners may have varied expertise on and subjective preferences of many topics. The reasons of which can be categorised as follows.

- **Diverse expertise**: learners (or students) may have varied levels of pre-knowledge and skills stemmed from their distinguished educational backgrounds and individual experiences. A typical example is that students from different departments surely have diverse expertise and domain knowledge.

- **Subjective preferences**: learners are likely to have subjective interests and preferences, which influence their desirable learning contexts to a great extent, e.g., some may prefer sport news while others enjoy science fictions.

The second sub-question is how to select learning tasks according to the constructed learner model. Research on this question has already been widely conducted in domains such as recommender systems (Li et al., 2013). In this article, we focus on the first sub-question from the perspective of how to model the preferred learning contexts of a language learner in a non-intrusive manner. This question is similar to the issue of how to obtain the preferred learning contexts of a learner in an implicit way according to his or her learning logs. Topic familiarity is an important factor here as tasks with contexts that are more familiar to a learner are found to lead to significantly better word learning performance (Pulido, 2003). It can therefore be inferred that recommending personalised word learning tasks with preferred or familiar learning contexts can promote effective word learning. For example, a cloze exercise with a preferred topic may stimulate the learning interest of learners. Therefore, in this article, we propose:

1. a framework for word learning systems to automatically obtain the context familiarity of each learner based on their logs, including their historical learning materials, testing results as well as the their writing assignments
2. a personalised approach to recommending word learning tasks according to individual learners’ context familiarity.

To verify the effectiveness and usefulness of the proposed framework, we conduct the experiments by inviting the real participants, and then comparing the performance among different learner groups.

The remaining sections of this article are structured as follows. In Section 2, related studies to this research are reviewed. We describe each sub-component of the proposed framework, which models the preferred learning contexts (context familiarity) of each learner based on the learning logs in Section 3. In Section 4, we present a personalised task recommendation approach to catering the personal preferred learning contexts. The experimental settings, procedures, marking criteria, results and implications are introduced in Section 5. Finally, we summarise this research and discuss the future research plans in Section 6.

2 Related work

In this section, we mainly review the related work and existing findings in two relevant areas: word learning, learner modelling and e-learning systems for language learning.
2.1 Word learning

Research in word learning can be generally summarised in three categories. One of them focuses mainly on word knowledge. It is commonly acknowledged that word knowledge is a continuum of one unique system containing both productive and receptive knowledge (Webb, 2005). Some researchers (Read and Chapelle, 2001; Nassaji, 2006) argue that there is a distinction to be made between the breath dimension and depth dimension of word knowledge, which is a model for measuring word knowledge. Particularly, the breadth dimension (a.k.a. vocabulary size) is the quantity of words acquired by learners at a specific level of language proficiency (Mehrpour et al., 2011), while the depth dimension refers to the quality of words known by a learner (Schmitt, 2008).

The other category concentrates on word learning process and facilitative factors for it. Fraser (1999) believes that word learning naturally is a cumulative process in an incremental way. Laufer and Hulstijn (2001) propose the involvement load hypothesis (ILH) to evaluate the effectiveness of diverse tasks in promoting incidental word learning. There are also many other studies (Hulstijn and Laufer, 2001; Williams, 2012; Godfroid et al., 2013) which attempt to verify the validity and reliability of this hypothesis.

2.2 Learner modelling

As the development of web-based learning systems from the late 1990s to early 2000s, researchers started to adopt the data mining and machine learning for learner modelling (Tang and McCalla, 2002; Webb et al., 2001). Later, Castillo et al. (2003) developed an adaptive incremental version of Naïve Bayes, to model a prediction task based on learning styles in the context of an adaptive hypermedia educational system. Brusilovsky et al. (2005) presented a generic student modelling server developed for a distributed e-learning architecture called knowledge tree. Brusilovsky and Millán (2007) discussed the student modelling along three dimensions: what is being modelled, how it is modelled, and how the models are maintained. By taking the teacher collaborations into account, Gaudioso et al. (2009) proposed a learner model to support the interaction among teachers and students. Hsiao et al. (2011) explored a social extension of open student modelling which can be visualised in terms of student progress. To keep student model update during the learning process, Jeremić et al. (2012) made use of a knowledge assessment method based on fuzzy rules. More recently, Lamb et al. (2014) considered the student cognitive processes by employing the computational model in the form of an artificial neural network (ANN) in a learner modelling. Xie et al. (2016) has built a learner model by taking the social media information and previous involvement loads of task for word learning.

2.3 E-learning systems for language learning

The era of big data witnesses rapid development and great popularity of e-learning systems (Li et al., 2009, 2013). Existing research basically follows the paradigm of intentional word learning rather than incidental word learning models (Zou et al., 2014). Loucky (2012) presents a task-based distance learning to optimise the vocabulary development of language learners. A blended learning environment named ‘ArabCAVL’ is developed by Essam (2010) to facilitate vocabulary acquisition of Arab students. Marc
et al. (2014) exploits the augmented reality (AR) techniques to enhance vocabulary learning and compare learning performance of various AR-based systems.

The popularity of mobile devices in recent years results in the ubiquitous word learning systems for learners. Through tracking users’ learning logs in mobile phones, Chen and Chung (2008) proposed a personalised ubiquitous system for English word learning according to the item response theory. Chen anf Li (2010) further improves their ubiquitous learning system by integrating the context-aware techniques which enable systems to be adapted according to learning contexts. From the perspective of learners, some studies focus on factors that positively facilitate word learning. Specifically, Basoglu and Akdemir (2010) compare the effectiveness of a mobile-based learning system and a game-based one. Huang et al. (2012) investigate the characteristics of ubiquitous word learning systems such as usefulness and ease-of-use, and find that active learners consider more about the feature of usefulness whereas inactive learners take into account more issues about ease-of-use.

3 The framework of identifying context familiarity

3.1 Problem formulation

The overall framework of identifying context familiarity can be constructed by measuring degrees of familiarity of each word for diverse learners from a collection of documents such as historical learning tasks and testing results related to the learner. Formally, we model the overall framework of identifying context familiarity as a mapping function $\theta$ between the set of documents $D$ and the set of learner profiles $L$ as follows.

$$\theta: D \rightarrow L$$  

where $L$ is in the form of word-weight values to indicate the familiarity of each word for learners, and each element $d \in D$ is essentially a document that can be modelled as a set of words $d = \{w \mid w \in d\}$.

Note that we do not limit $D$ for learning documents, which only mirror the familiarity of a learner in an objective way. For example, test results reflect whether a learner understands meanings of target word or not. However, we believe that self-motivation driven by individual preferences may also have positive impacts on word acquisition. We therefore invite learners to complete questionnaires to indicate their subjective preferences on contexts (topics) for learning. These questionnaires are also included in the set $D$. To distinguish them from other learning documents, we notate learning documents and questionnaires as $D'$ (objective documents) and $D''$ (subjective documents) respectively, where $D' \cup D'' = D$, $D' \cap D'' = \emptyset$. The detail definitions and corresponding process of set $D$ (including $D'$ and $D''$) and set $L$ are introduced in the remaining parts of this section.

3.2 Learner profile with familiarity

In this sub-section, the definition of learner profile is introduced. As we believe that the degree of learners’ familiarity with learning contexts have positively effects on vocabulary learning, the degree of learners’ familiarity with various learning contexts is included in the learner profile. However, it is impossible to include all contexts in a
Modelling second language learners for learning task recommendation

learner profile because the learning contexts are naturally permutation of all words. The quantity of learning contexts are \( \sum_{i=1}^{n} P^{i} \), where \( P \) is the permutation, \( |V| \) is the vocabulary size, and \( i \) is the length of the context. To address this issue, we include the familiarity of each word rather than all possible learning contexts in the learner profile.

**Definition 1:** let \( \{w_1, w_2, \ldots, w_n\} \in V \) and \( \{e_1, e_2, \ldots, e_n\} \) be the corpus of all words and the corresponding degree of familiarity of each word for learner \( l_i \), the learner profile of \( l_i \) is denoted by a vector \( l_i \) as:

\[
\bar{l}_i = (w_1, e_1^i; w_2, e_2^i; \ldots; w_n, e_n^i)
\]

As learning contexts are basically consisted of words, the degree of familiarity of a learning context can be regarded as the expectation (i.e., the weighted mean) of the familiarity degrees of all words that form this context. Given a learning context \( c = \{w_1', q_2', w_m'\} \), the degree of familiarity for learner \( l_i \) therefore can be calculated as follows.

\[
f(c, \bar{l}_i) = \sum_{j=1}^{m} r(w_j') \cdot e_j^i
\]

where \( f(c, \bar{l}_i) \) is the function of calculating familiarity degree of a learning context for a learner, and \( r(w_j') \) is the ratio of \( w_j' \) appearing in the context \( c \). Note that words in the context \( c \) is a sequence and there may be duplicated words. To eliminate the negative influence of useless high frequency words (e.g., ‘the’, ‘an’) we pre-process learner profiles, learning contexts as well as other relevant documents by deleting all words that are in the stop-words list created by Google (Google Stop Words List, 2014).

### 3.3 Objective vocabulary familiarity

The objective context familiarity refers to the context familiarity obtained from objective documents (i.e., learning documents \( D' \)). Specifically speaking, we mainly use three kinds of learning documents: texts for reading comprehension (denoted as \( D' \)), writing assignments for short essays (denoted as \( D' \)) and test papers for word learning practice in the format of multiple choices (denoted as \( D' \)).

#### 3.3.1 Learning essays

For each article \( (d' \in D') \), it can be represented by a bag-of-words. We hypothesise that the degree of familiarity is positively correlated with frequencies of all words in an essay. Therefore, we employ the model of term frequency and inverse document frequency (TF-IDF) to calculate the familiarity (Manning et al., 2008).

\[
f_i (d'_i, w_s) = (1+\log n(w_s)) \times \log \left(1 + \frac{N_i}{N_i(w_s)}\right)
\]

where \( n(t) \) denotes the frequency of word \( w_s \) appearing in an essay \( d'_i \) learnt by learner \( i \), \( 1 + \log n(w_s) \) is the log normalisation, \( N_i \) is the total number of essays learnt by this learner, and \( N_i(w_s) \) is the total number of essays containing word \( w_s \). The sum of degrees
of familiarity for each word to a learner is the cumulation of all learning essays with an upper limit ‘1’.

3.3.2 Writing assignments

For writing essays, it can also be denoted by a bag-of-words. However, in addition to the pre-process step of deleting all stop-words, the words with errors and typos should not be taken into account while meaning the degree of familiarity. Therefore, we parse each word with WordNet and check whether the word exists in the WordNet or not (WordNet, 2014). Rather than employing the TF-IDF model, we believe that writing process involves using words that have been already stored in a learner’s memory. The ratio of the use of words reflects the degree of familiarity. Therefore, the quantity of writing essays is a significant factor, and we adopt the ratio to represent the degree of familiarity.

\[ f_i(D_r, w_i) = \frac{n(w_i)}{N(D_r)} \]  \hspace{1cm} (4)

where \( n(w_i) \) is the total frequencies of word \( w_i \) among all writing essays, and \( N(D_r) \) is the total number of words in all writing essays.

3.3.3 Test papers

As the main purpose of a test paper is to exam whether a student acquires the knowledge of target words or not, we can interpret the ratio of correct answers of a target word among all tests as the degree of familiarity.

\[ f_i(D_r, w_i) = \frac{c(w_i)}{N(w_i)} \]  \hspace{1cm} (5)

where \( c(w_i) \) is the number of correct answers for the target word \( w_i \), \( N(w_i) \) is the total number of tests of the word \( w_i \). Note that for those test papers which do not show detail test results for each target words, we take the overall test score as the degree of familiarity.

Through converting the above three kinds of documents into the degree of familiarity, we can obtain the overall objective context familiarity. However, there are still two minor issues to be addressed, which are:

1. how to model the temporal signals of objective documents

2. how to assign weights to three kinds of familiarity from three data sources during the aggregation.

To solve the first issue, we employ the forgetting curve (Wixted and Ebbesen, 1997) to incorporate the temporal signal of each related document. The basic idea is that tasks being learnt recently should contribute more to the retention of target words. Formally, we use the following time-decayed function:

\[ f^* = e^{-\Delta t/\tau} \]  \hspace{1cm} (6)

where \( \Delta t \) is the time interval between the current moment and when the related document has been completed, \( f \) is abstract function for equations (3) (4) and (5). Note that we need to embed this function during the process of calculating the overall degree of familiarity.
[e.g., equations (4) and (5)]. For simplicity, we do not show how to embed equation (6) to these functions.

To tackle the second issue, we check three kinds of documents which correspond to three incidental word learning tasks (i.e., the tasks of reading comprehension, essay writing and cloze test). As suggested by the involvement load hypothesis (hereafter, ILH) (Laufer and Hulstijn, 2001), the task of essay writing induces strong evaluation, the degree of prominence of which can be marked as two pluses (evaluation, ++), and the task of reading comprehension and cloze induces moderate evaluation marked by one plus (evaluation, +). Therefore, we adopt the following aggregation method to obtain the overall objective context familiarity:

\[ f_o = \alpha_1 \cdot f_e + \alpha_2 \cdot f_s + \alpha_3 \cdot f_i \]  

(7)

where \( \alpha_1 = \alpha_3 = 1/4, \alpha_2 = 1/2 \) indicate their involvement loads induced in the aggregation process. Note that \( f_o \) is a simplified notation of \( f_o(w_i') \).

### 3.4 Subjective vocabulary familiarity

We also believe that individual preferences for the contexts facilitate word learning. For example, a football fan is likely to be more willing to learn words from the report about FIFA world cup than a research article on the new breakthrough in quantum mechanics. However, it is infeasible to specify the subjective preferences for all target words by each learner. Therefore, we invite learners to complete questionnaires to indicate their subjective preferences to contexts (topics) for learning. The questionnaire includes all pre-defined topics associated with some typical words (e.g., topic: food, words: bread, chip, steak, etc). The learners are required to give a score from ‘strongly dislike’ to ‘strongly like’ (ranging from ‘1’ to ‘5’) for each topic.

A problem here is how to assign the subjective familiarity to each word when you know the individual preferences for a topic. The solution is that we adopt the latent Dirichlet allocation (LDA) to identify topics and the associated typical words (Blei et al., 2003). For each word, we have probability distribution \( p(w_t|t) \) over all topics. Next, we use the expectation to denote the subjective familiarity as follows.

\[ f_i(w_i') = \sum_j s_i(t_j) \times p(w_i|t_j) \]  

(8)

where \( s_i(t_j) \) is a score given by learner \( i \) to topic \( t \) (note that the score is normalised to the scale of [0, 1]), \( p(w_i|t_j) \) is the probability distribution of \( w_i \) for a topic \( t_j \), and \( f_i(w_i') \) denotes the subjective familiarity for word \( w_i \) to learner \( i \).

Therefore, we can obtain the final familiarity of each word in learner profile (as defined in Definition 1) by aggregating the objective and subjective familiarity as follows.

\[ \varepsilon_i = \beta_1 f_o(w_i') + \beta_2 f_s(w_i') \]  

(9)

where two parameters \( \beta_1 \) and \( \beta_2 \) are to adjust the weight of two kinds of familiarity, and we adopt the optimal combinations (\( \beta_1 = 0.6 \) and \( \beta_2 = 0.4 \)) suggested by Cai et al. (2010).
In this section, we introduce how to recommend incidental word learning tasks based on the familiarity-based learner profile obtained in Section 3. The learning context associated with a task can also be represented by a bag-of-words paradigm. Formally, we define the task profile to denote the learning context as follows.

**Definition 2:** let \( \{w_1, w_2, \ldots, w_n\} \in V \) and \( \{\delta^e_1, \delta^e_2, \ldots, \delta^e_n\} \) be the corpus of all words and the corresponding degree of relevance to a learning context of task \( t_a \). The task profile of \( t_a \) is denoted by a vector \( \tilde{t}_a \) as:

\[
\tilde{t}_a = (w_1, \delta^e_1; w_2, \delta^e_2; \ldots; w_n, \delta^e_n)
\]

where the degree of relevance is the ratio of the word appearing in the learning context of the learning task \( t_a \).

To recommend incidental word learning tasks with more familiar contexts to learners, it is essential to employ a reasonable measurement to estimate how familiar the learning context is to the learner when the task profile and the learner profile are provided. Research on profile-based information retrieval (IR) has found that the conventional measurement in IR, for instance cosine similarity, may be unsuitable due to the fact that the nature of the problem is to find the most familiar task profile rather than the most similar one (Cai et al., 2010; Xie et al., 2012). Therefore, in this research, we adopt the projection operation from the learner profile and the task profile as the measurement of the degree of familiarity.

\[
s(\tilde{t}_a, \tilde{l}_t) = \|\tilde{l}_t\| \cos \alpha
\]

where \( \|\tilde{l}_t\| \) is the Euclidean length of vector \( \tilde{l}_t = <e'_1, e'_2, \ldots, e'_d> \), \( \alpha \) is the angle between two vectors. The function \( s \) here is to project learner profile to the task profile.

The motivation of using projection is that the degree of familiarity can be interpreted as the question of how familiar each word in the learning context is to a learner. To answer the question, the learner profile is therefore projected to the task profile (i.e., the learning context) to measure the holistic degree of familiarity. In sum, we recommend the task with the highest degree of familiarity [as calculated by equation (10)] to the learner.

\[
t^* = \arg \max_{t \in T} s(\tilde{t}_a, \tilde{l}_t)
\]

where \( T \) is a set of tasks available in the system for the same target words, yet the task \( t^* \) with the highest degree of familiarity is recommended to the learner. As noted by previous studies on personalised word learning task recommendation (Zou et al., 2014), other aspects such as task diversity or word coverage may also be conducive to word learning. In fact, these aspects can obviously improve the learning effectiveness if they are reasonably and suitably integrated with the context familiarity. In this work, we focus on the issue of how to exploit the context familiarity and keep the integration as our future work.
5 Experiments

In this section, we describe the experiment conducted to evaluate the proposed approach to personalised recommendations. After introducing the details of the materials and subjects in the experiment, we present the marking criteria (metric) we applied to evaluate the effectiveness of subjects’ learning of target words. The detail experiment processes are also explained. We then report the experimental results and discuss the underlying implications of our findings for designing both pedagogical activities and word learning systems.

5.1 Materials and subjects

A pilot study was conducted to select the set of target words for our experiment. As the participants were freshmen from a university in Hong Kong, their language proficiency levels are normally in the range of level 3 to 5 in the HKDSE English Language Subject, which are corresponded to scores from 5.48 to 6.99 in the International English Language Testing System (HKDSE, 2015). Results of the standard vocabulary tests (Vocabulary Size Test, 2015), which was conducted in the pilot study, show that the vocabulary sizes of our participants are in the range of 6,000 to 8,000 words.

Therefore, we selected ten words from the most frequently used 9,000 to 14,000 words based on the vocabulary levels. The results of our pilot study showed that these words were unlikely to be familiar to the participants of the experiment. These target words were embedded in the texts of three categories of learning contexts, namely information technology, science and literature. Adapted from fictions, academic articles and news reports, the learning contexts were further polished or re-written by three native speakers so that all participants are able to understand their general ideas literally.

Table 1 shows an example of three different learning contexts for the target word ‘ubiquitous’.

Table 1 An example of three different learning contexts for the target word ‘ubiquitous’

<table>
<thead>
<tr>
<th>Categories</th>
<th>Learning contexts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information technology</td>
<td><em>Ubiquitous</em> computing is a paradigm in which the processing of information is linked with each activity or object as encountered.</td>
</tr>
<tr>
<td>Science</td>
<td>The advantages of Coca Cola as an extracted liquid are its <em>ubiquitous</em> availability and readiness for use.</td>
</tr>
<tr>
<td>Literature</td>
<td><em>Ubiquitous</em> weeds are laid in the wall of this ancient castle.</td>
</tr>
</tbody>
</table>

Table 2 Descriptive statistics of the subjects

<table>
<thead>
<tr>
<th>Group type</th>
<th>Group name</th>
<th>#subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>37</td>
</tr>
<tr>
<td>Age</td>
<td>17–19</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>19–21</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>21–23</td>
<td>3</td>
</tr>
<tr>
<td>Region</td>
<td>Hong Kong</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>Other countries</td>
<td>8</td>
</tr>
</tbody>
</table>
The 82 participants were freshmen from a university in Hong Kong. They were randomly divided into four groups in the experiment. The detailed descriptive statistical information about the demographical attributes of the participants is shown in Table 2.

5.2 Marking criteria

We adopt a modified vocabulary knowledge scale (MVKS), which has been proposed by Folse (2006). The main reason of employing the MVKS in our research is that it emphasises the measurement of both subjects’ receptive and productive knowledge of target words, which are learnt by subjects through incidental word learning. Specifically speaking, MVKS is a three-point scale as shown in Table 3. The scoring system developed by Hulstijn and Laufer (2001) was applied.

Table 3  Folse’s (2006) modified vocabulary knowledge scale

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I do not know what this word means.</td>
</tr>
<tr>
<td>2</td>
<td>I know this word. It means _______________________________________. (provide an English synonym or a translation in your native language</td>
</tr>
<tr>
<td>3</td>
<td>I can use this word in a good example. Write your sentence here: ____________________________________________ (if you do #3, you must do #2 also.)</td>
</tr>
</tbody>
</table>

Subjects’ answers to meanings of target words were graded in this way:
1. a score of ‘0’ will be given to an answer with completely incorrect meaning
2. a score of ‘2.5’ will be given to an acceptable equivalent of the target word
3. a score of ‘5’ will be given to a comparable meaning of the target word.

Yet subjects’ original sentences by using target words were graded in this way:
1. a score of ‘0’ will be given to a sentence with completely improper semantic context for the target word
2. a score of ‘2.5’ will be given to a sentence with proper semantic context but ungrammatical usage
3. a score of ‘5’ will be given to a sentence with proper semantic context and grammatical usage.

5.3 Experimental processes and results

In this sub-section, we present the detail experimental processes and report the corresponding results. More specifically, the overall experimental processes include the following four detail stages.

a  Objective familiarity acquisition. 82 freshmen were enrolled in an English course with a period of 13 weeks. Consented by the subjects, we collected their assignments, learning essays as well as testing papers for this course from the MOODLE e-learning system. Averagely, there were 4.4 assignments, 6.8 learning essays and 2.9 testing papers for each student to acquire their objective vocabulary familiarity according to the Equations mentioned in Section 3.3. Based on these
learning documents, the sizes of vocabulary of the 82 learners were from 823 to 1,214 words by eliminating the stop words.

b  **Subjective familiarity acquisition.** We employ a corpus including 143 essays in information technology, science and literature to tune the parameters in the topic model. The number of topics was specified as three and the Gibbs sampling has been adopted. According to the preference scores of the subjects, which were evaluated through an online survey by asking the participants to indicate their preferences of the three topics in a five-scale system, 1 representing ‘not like it at all’ and 5 representing ‘like it very much’), the subjective vocabulary familiarity of each subject has been also obtained and aggregated with objective vocabulary familiarity as introduced in Section 3.4.

c  **Personalised task recommendation.** We divided 82 subjects into three groups, the details of which are summarised in Table 4. Group A was the control group. The 27 participants in this group were provided with randomly assigned learning tasks without taking into account their topic familiarity. The tasks recommended to the participants in group B were selected based on the participants’ objective familiarity with the learning contexts. Their subjective vocabulary familiarity was not considered. The learning contexts provided to the 28 subjects in group C were selected according to their aggregated vocabulary familiarity, including both objective and subjective vocabulary familiarity.

d  **Word learning evaluation.** The immediate post-tests were conducted right after a 30 minutes learning period. One week later, these subjects were invited to take a delayed post-test. Both immediate and delayed post-tests applied the same format to examine the subjects’ knowledge of target words as illustrated in Table 3. Three language experts were invited to mark the immediate and delayed post-tests according to the marking criteria introduced in Section 5.2. In case of any inconsistency of marking, the medium of the scores by three experts was used as the final score.

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Three groups of 82 subjects in the experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group</strong></td>
<td><strong>N</strong></td>
</tr>
<tr>
<td>A</td>
<td>27</td>
</tr>
<tr>
<td>B</td>
<td>27</td>
</tr>
<tr>
<td>C</td>
<td>28</td>
</tr>
</tbody>
</table>

The results of immediate and delayed post-tests for the three groups are shown in Figure 1. Group C achieves the best performance in both immediate and delayed post-tests among all groups, while groups B and A are the first and second runner-up in the two post-tests. Independent sample t-tests were conducted at the significance level was achieved ($p < 0.05$). Such results support the rationale of the proposed framework. That is, the participants’ learning performance is better when more kinds of vocabulary familiarity are taken into account. Note that the differences of the scores of the three groups of participants were smaller in the immediate post-test (61.4, 64.2 and 68.7) than the delayed post-test (43.7, 44.8 and 45.4). However, the differences were still significant.
as shown by the results of the independent sample t-tests. This is consistent with the findings of previous work (Wixted and Ebbesen, 1997) that the word retention rate would drop to a similar level without any review after a certain period.

**Figure 1** The results of immediate and delayed post-tests of three groups

5.4 Learning enjoyment survey

To evaluate the participants’ subjective perceptions of the learning processes, we also conducted a survey with five rating scales. The rating score 1 means ‘strongly disagree’ while 5 denotes ‘strongly agree’. The questionnaires mainly contain two aspects:

1. ‘the learning contexts in the tasks are very familiar to me’ (denoted as ‘familiarity’)
2. ‘the learning contexts in the tasks are very interesting to me’ (denoted as ‘interestingness’).

The group means are shown in Figure 2. The trends among the three groups were similar to those shown in Figure 1. Group C had the highest scores in the two aspects of familiarity and interestingness, while group A had the lowest scores in both aspects. We may conclude that the integration of vocabulary familiarity in a learner model can increase the learning interest and performance as better learning results from more familiar contexts. Another interesting finding demonstrated in Figure 2 is that the increase of interest between group A and group B (i.e., 2.37 – 2.22 = 0.15) is much less than the increase of familiarity (i.e., 2.81 – 2.33 = 0.48). A possible explanation of this is that different from group A, group B involved only objective vocabulary familiarity but not subjective vocabulary familiarity; and without considering the subjective input of the preferred topics, learners may still not like the learning contexts. Better learning performance was observed in Group C when the subjective vocabulary familiarity was taken into account, as both the scores of familiarity and interestingness were improved and reached same levels.
5.5 Implications

This research may shed lights on the design of pedagogical activities in classroom and the development of e-learning systems. We detail the implications in these two areas as follows.

- **Pedagogical implications.** Lecturers are suggested to encourage students to indicate their preferences on essays, videos and other types of materials to be learnt in class rather than use a preset textbook. The learning preferences and pre-knowledge of students should also be carefully considered while developing class activities, e.g., the selection of favourable topics for a writing assignment. In addition, the involvement of students can increase their sense of ownership of the class as well as the motivation of learning.

- **Implications for e-learning systems.** E-learning systems for vocabulary acquisition are advised to provide more options for users to select their preferable learning contexts. For example, let users select their favourite topics such as sports, literature or financial news through a dropdown list of essays with the same target words. Those words that have been learnt can be also included in the learning contexts for consolidation while learning other new target words. These implications may stimulate our future research.

6 Conclusions and future work

In this research, we study the research question of how to model the preferred learning contexts for individual learners according to their historical learning logs. The issues of how to model the subjective and objective context (vocabulary) familiarity and how to exploit the context familiarity for task recommendations are also discussed. We have found that the three typical learning documents: learning essays, writing assignments, and
test papers can reflect the preferred learning contexts (or the context familiarity) of individual learners effectively. Moreover, the extensive experiment which involves 82 subjects has shown the effectiveness of the proposed framework including the learner model and the recommendation approach. In addition, we discussed the implications of this research for the design of pedagogical activities and e-learning systems for vocabulary acquisition.

We plan to investigate the following directions in future studies:

1. as the different learning contexts for the same target words were manually pre-defined in the system, the processes of which are tedious and time-consuming, the issue of how to automatically extract the learning contexts from a corpus or web will be investigated

2. how to integrate the words that have been learned recently into the materials that will be learned later so as to consolidate the learners’ knowledge of the words is another interesting topic.

Acknowledgements

The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/FDS11/E06/14) and the start-up research grant (RG 37/2016-2017R) of The Education University of Hong Kong.

References


Modelling second language learners for learning task recommendation


