Ogata, H. et al. (Eds.) (2015). Proceedings of the 23rd International Conference on Computers in Education. China: Asia-Pacific Society for Computers in Education

Identifying Context Familiarity for Incidental Word Learning Task Recommendations

Haoran XIE[a], Di ZOU[b*], Fu Lee WANG[a], Tak-Lam WONG[c] & Qingyuan Wu[d]
[a]*Caritas Institute of Higher Education, Hong Kong SAR*
[b]*English Language Centre, The Hong Kong Polytechnic University, Hong Kong SAR*
[c]*Department of Mathematics and Information Technology,*
The Hong Kong Institute of Education, Hong Kong SAR
[d]*School of Management, Beijing Normal University, Zhuhai, China*
* daisyzou@polyu.edu.hk

Abstract: Incidental word learning tasks are widely adopted in pedagogical activities and self-paced learning processes, given their advantages of providing rich contexts and training on other language skills. While these tasks, diverse learners are often provided with the same contexts. For example, a cloze test with the same essay may be offered to all users in an e-learning system for learning target words. However, different learners may have varied expertise on and subjective preferences of many topics. Hence the provided unified learning context may be unfamiliar to some learners. The learning effectiveness is therefore likely to be negatively influenced. In response to a call to solve this problem, we propose in this paper a framework for word learning systems to automatically identify the context familiarity of individual learners based on their logs. A personalized approach to accurate recommendations of incidental word learning tasks is also devised according to the individual context familiarity. The results of our experimental studies on real participants show that the proposed framework and method promote significantly more effective word learning and increase the learning enjoyment greatly than conventional approaches with unified learning contexts.

Keywords: Incidental word learning, context familiarity, task recommendation, learner profile, e-learning

1. Introduction

Incidental word learning tasks are widely adopted in pedagogical activities and self-paced learning processes. This is mainly because incidental word learning provides not only richer contexts but also more opportunities of practicing on other language skills such as reading, listening, speaking and writing, compared with intentional word learning (Zou et al., 2014). While doing incidental word learning tasks, diverse learners are often provided with the same contexts. For example, a cloze test with the same essay may be offered to all users in an e-learning system for learning target words.

However, the provided unified learning context is perhaps unfamiliar to some learners, the reasons of which can be categorized 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 learning effectiveness therefore can be negatively influenced if the unified learning contexts are offered to all users. To tackle this problem, we therefore propose (i) 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; and (ii) design a personalized approach to recommendations of incidental word learning tasks according to the individual context familiarity. Our hypothesis is that those tasks with contexts that are more familiar and/or engaging promote significantly more effective word learning. This is partially supported by a study on topic familiarity and incidental word learning (Pulido, 2003). To further explore the effectiveness of the proposed framework and approach in e-learning systems, we invite some subjects to participate in the experiment. The results of our research show that the proposed framework and method can promote significantly more effective word learning compared to conventional approaches with unified learning contexts.

The reminder of this paper is organized as follows. In Section 2, we review the related work and findings from past research. The framework which is designed to automatically identify context familiarity of each learner based on their logs is introduced in Section 3. In Section 4, we present the personalized approach to recommendations of incidental word learning task according to the individual context familiarity. The experimental settings, procedures as well as the results are reported in Section 5. Finally, we summarize this research and discuss potential issues to be further explored.

2. Related Work

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

2.1 Incidental Word Learning

Research in incidental word learning can be generally summarized in two 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 & Chapelle 2001; Nassaji, 2006) argue that there is a distinction to be made between the breadth 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 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; Li et al., 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 optimize 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 phone, Chen and Chung (2008) proposed a personalized ubiquitous system for English word learning according to the item response theory. Chen et al., (2010) further improves their ubiquitous learning system by integrating the context-aware techniques which enable systems to be adapted according to learning contexts.

3. The Framework of Identifying Context Familiarity

3.1 Problem Formulation

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

$$\theta\text{:}\ \ \mathrm{D} \rightarrow \mathrm{L} \tag{1}$$

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 modeled as a set of words $d = \{w \mid w \in d\}$.

3.2 Learner Profile with Familiarity

In this subsection, the definition of the learner profile is introduced. As we believe that the degree of familiarity of learning contexts have positively effects on vocabulary learning, the degree of familiarity for various learning contexts is included in the learner profile. However, it is impossible to include all contexts in a learner profile because the learning contexts are naturally permutation of all words. The quantity of learning contexts are $\sum_{i=1}^{n} \mathrm{P}_i^{|V|}$, 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 $\{\varepsilon_1^i, \varepsilon_2^i, \ldots \varepsilon_n^i\}$ 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 $\vec{l}_i$ as:

$$\vec{l}_i = (w_1, \varepsilon_1^i; w_2, \varepsilon_2^i; \ldots w_n, \varepsilon_n^i)$$

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 $c = \{w_1^{'}, w_2^{'}, \ldots w_m^{'}\}$, the degree of familiarity for learner l_i therefore can be calculated as follows.

$$f(c, \vec{l}_i) = \sum_{j=1}^{m} r(w_j^{'}) \cdot \varepsilon_j^i \tag{2}$$

where $f(c, \vec{l}_i)$ 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 may contain 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, 2014).

3.3 Objective Vocabulary Familiarity

The objective context familiarity refers to the context familiarity obtaining from objective documents (i.e., learning documents D^{o}). Specifically speaking, we mainly use three kinds of learning documents: learning articles for reading comprehension (denoted as D^{r}), writing assignments for short essays (denoted as D^{e}) and test papers for word learning practice in form of multiple choices (denoted as D^{t}).

Learning Essays. For each article ($d^{r} \in D^{r}$), it can be represented by a bag-of-words. We hypothesize 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, 2008).

$$f_r(\mathrm{d}_i^r, w_x) = (1 + \log n(w_x)) \times \log(1 + N_i / N_i(w_x)) \tag{3}$$

where $n(t)$ denotes the frequency of word w_x appearing in an essay d_i^r learnt by learner i, 1+log$n(w_x)$ is the log normalization, N_i is the total number of essays learnt by this learner, and $N_i(w_x)$ is the total number of essays containing word w_x. The sum of degrees of familiarity for each word to a learner is the cumulation of all learning essays with an upper limit “1”.

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.

$$f_e(\mathrm{D}_i^e, w_x) = n(w_x) / N(\mathrm{D}_i^e) \tag{4}$$

where $n(w_x)$ is the total frequencies of word w_x among all writing essays, and $N(\mathrm{D}_i^e)$ is the total number of words in all writing essays.

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.

$$f_t(\mathrm{D}_i^t, w_x) = c(w_x) / N(w_x) \tag{5}$$

where $c(w_x)$ is the number of correct answers for the target word w_x, $N(w_x)$ is the total number of tests of the word w_x. 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. To assign weights to three kinds of familiarity from three data sources during the aggregation, we adopt the ILH theory (Laufer and Hulstijn 2001). As suggested by the ILH theory (Laufer and Hulstijn 2001), we can assign the essay writing task with two involvement loads, while assign one to the task of reading comprehension and cloze. Therefore, we adopt the following aggregation method to obtain the overall objective context familiarity:

$$f_o = \alpha_1 \cdot f_r + \alpha_2 \cdot f_e + \alpha_3 \cdot f_t \tag{6}$$

where α_1=α_3=1/4, α_2= 1/2 to indicate their loads suggested by the ILH theory during their aggregation process. Note that f_o is a simplified notation of $f_o(w_x^i)$.

3.4 Subjective Vocabulary Familiarity

We also believe that individual preferences for the contexts may also positively facilitate word learning. Thus, 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 $p(w|t)$ over all topics. Next, we use the expectation to denote the subjective familiarity as follows.

$$f_s(w_x^i) = \sum_j s_i(t_j) \times p(w_x \mid t_j) \tag{7}$$

where $s_i(t_j)$ is a score given by learner i to topic t (note that the score is normalized to the scale of [0, 1]), $p(w_x|t_i)$ is the probability distribution of w_x for a topic t_j, and $f_s(w_x^i)$ denotes the subjective familiarity for word w_x to learner l_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.

$$\varepsilon_x^i = \beta_1 f_o(w_x^i) + \beta_2 f_s(w_x^i) \tag{8}$$

where two parameters β_1 and β_2 are to adjust the weight of two kinds of familiarity, and we adopt the optimal combinations (β_1 = 0.6 and β_2 = 0.4) suggested by Cai et al. (2010).

4. Personalized Task Recommendation

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 $\{ \delta_1^a,\ \delta_2^a, \ldots \delta_1^a \}$ 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 $\vec{t}_a$ as:

$$\vec{t}_a = (w_1, \delta_1^a; w_2, \delta_2^a; \ldots w_n, \delta_n^a)$$

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.

$$s(\vec{t}_a, \vec{l}_i) = \| \vec{l}_i \| \cdot \cos\alpha \tag{9}$$

where $\| \vec{l}_i \|$ is the Euclidean length of vector $\vec{l}_i = < \varepsilon_1^i, \varepsilon_2^i, \ldots \varepsilon_n^i >$, α is the angle between two vectors. Essentially speaking, 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 to the learner.

$$t^* = \arg\max_{t_a \in T} s(\vec{t}_a, \vec{l}_i) \tag{10}$$

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.

5. Experiments

In this section, we describe the experiment conducted to evaluate the proposed approach to personalized recommendations. After introducing the details of the materials and subjects in the experiment, we then report the experimental results.

5.1 Materials and Subjects

A pilot study was conducted to determine 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 6000 to 8000 words.

Therefore, we selected 10 words from with the vocabulary levels of 9000 to 14000 and further verified that these words were not acquainted with participants in the experiment. These words were embedded into 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 can understand their general ideas literally. As mentioned, the participants are freshmen from a university in Hong Kong. There are totally 82 subjects who are further divided into 3 groups in the experiments.

5.2 Experimental Results

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 Group

B and A are the first and second runner-up in the two post-tests. The differences among groups reach at significance level according to t-test analysis ($p<0.05$). Such results support the rationale of the proposed framework. That is, the more varied kinds of vocabulary familiarity proposed in our framework been consider, the more effective the learning contexts being recommended would be. Note that the differences among three groups in the immediate post-tests (61.4, 64.2 and 68.7) were smaller in the delayed post-tests (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.

Table 4: Three groups of 82 subjects in the experiments

Group	N	Descriptions
A	27	Control group without providing any familiarity-based contexts.
B	27	Contexts are recommended based on objective vocabulary familiarity (Eq. (6)).
C	28	Contexts are recommended based on aggregated vocabulary familiarity (Eq. (8)).

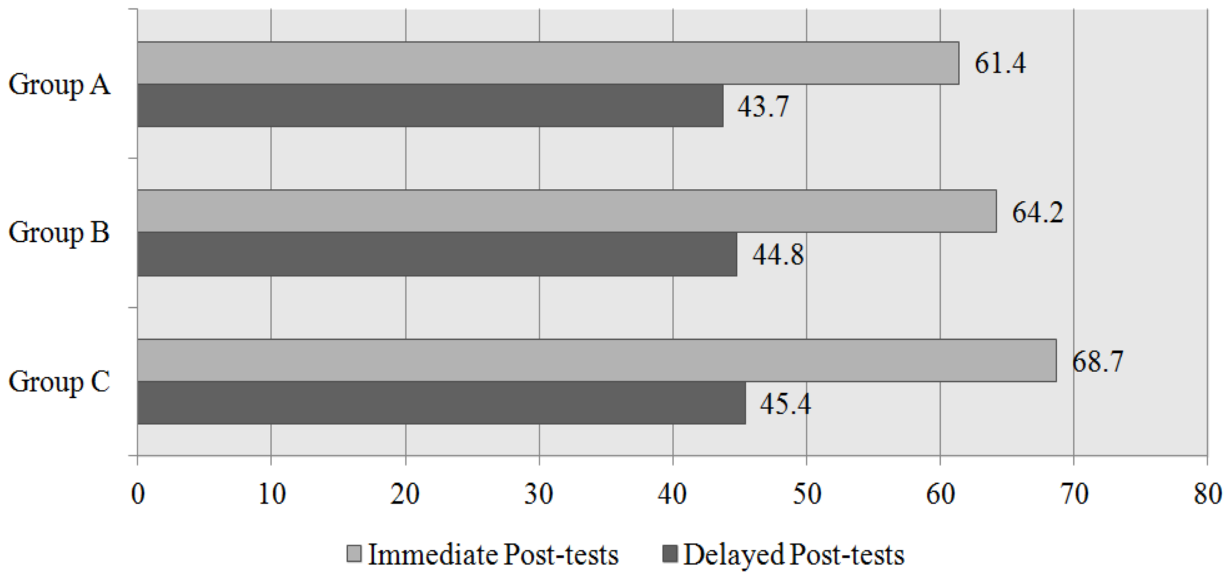

Figure 1. The results of immediate and delayed post-tests of three groups.

6. Conclusion and Future Work

In this research, we studied issues concerning how to identify the context familiarity from historical learning documents of learners, how to model the subjective and objective context (vocabulary) familiarity, as well as how to exploit the context familiarity and recommend preferable learning contexts to assist learners' vocabulary acquisition. We have also conducted an extensive experiment involving 82 subjects, the results of which verify the rationale of the proposed framework of context familiarity and personalized approach to recommendations of incidental word learning tasks according to the individual context familiarity. Furthermore, we discussed the implications for both the design of pedagogical activities and e-learning systems for vocabulary acquisition. For our future study, we plan to integrate those words that have been learnt recently into the upcoming learning contexts is another interesting topic as highlighted in the above systematical implications.

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).

References

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. the Journal of machine Learning research, 3, 993-1022.

Cai, Y., Li, Q., Xie, H., & Yu, L. (2010). Personalized resource search by tag-based user profile and resource profile. In Web Information Systems Engineering–WISE 2010 (pp. 510-523). Springer Berlin Heidelberg.

Chen, C. M., & Chung, C. J. (2008). Personalized mobile English vocabulary learning system based on item response theory and learning memory cycle. *Computers & Education*, 51(2), 624-645.

Chen, C. M., & Li, Y. L. (2010). Personalised context-aware ubiquitous learning system for supporting effective English vocabulary learning. *Interactive Learning Environments*, 18(4), 341-364.

Essam, R. (2010). Software Application for Computer Aided Vocabulary Learning in a Blended Learning Environment. *Master Thesis*, American University in Cairo, Cairo, Egypt.

Fraser, C. A. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in second language acquisition*, 21(02), 225-241.

Godfroid, A., Boers, F., & Housen, A. (2013). AN EYE FOR WORDS. *Studies in Second Language Acquisition*, 35(03), 483-517.

Google Stop Words List. (2015, March 25). Retrieved from https://code.google.com/p/stop-words/

Hong Kong Diploma of Secondary Education Examination (HKDSE) to IELTs. (2015, March 25). Retrieved from http://www.hkeaa.edu.hk/en/recognition/benchmarking/hkdse/ielts/

Horwitz, E. K. (1988). The beliefs about language learning of beginning university foreign language students. The *Modern Language Journal*, 72(3), 283-294.

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language learning*, 51(3), 539-558.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied linguistics*, 22(1), 1-26.

Li, Q., Lau, R. W., Wah, B., Ashman, H., Leung, E. W., Li, F., & Lee, V. (2009). Guest Editors' Introduction: Emerging Internet Technologies for E-Learning. *Internet Computing*, IEEE, 13(4), 11-17.

Li, Q., Luo, X., Wenyin, L., & Conati, C. (2013). Introduction to the special section on intelligent tutoring and coaching systems. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 4(2), 28.

Loucky, J. P. (2012). Designing distance learning tasks to help maximize vocabulary development. International *Journal of Virtual and Personal Learning Environments (IJVPLE)*, 3(2), 35-58.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to information retrieval (Vol. 1). Cambridge: Cambridge university press.

Marc, E. S., Arno, W. L., Takafumi, T., Goshiro, Y., Ma., T. R., Christian., S., & Hirokazu., K. (2014). Evaluating Augmented Reality for Situated Vocabulary Learning. *In Proceedings of 22nd International Conference on Computers in Education*, Nara, Japan, pp. 701-710.

Mehrpour, S., Razmjoo, S. A., & Kian, P. (2011). The relationship between depth and breadth of vocabulary knowledge and reading comprehension among Iranian EFL learners. *Journal of English Language Teaching and Learning*, 222 (53), 97-127..

Nassaji, H. (2006). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Modern Language Journal*, 90(3), 387-401.

Pulido, D. (2003). Modeling the role of second language proficiency and topic familiarity in second language incidental vocabulary acquisition through reading. *Language learning*, 53(2), 233-284.

Read, J., & Chapelle, C. A. (2001). A framework for second language vocabulary assessment. *Language Testing*, 18(1), 1-32.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language teaching research*, 12(3), 329-363.

Vocabulary Size Test. (2015, March 25). Retrieved from http://www.lextutor.ca/tests/levels/recognition/1_14k/

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27(01), 33-52.

Williams, J. (2012). The potential role (s) of writing in second language development. *Journal of Second Language Writing*, 21(4), 321-331.

Wixted, J. T., & Ebbesen, E. B. (1997). Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. Memory & cognition, 25(5), 731-739.

WordNet. (2015, March 25). Retrieved from http://wordnet.princeton.edu/

Xie, H., Qing, L., and Yi, C. "Community-aware resource profiling for personalized search in folksonomy." Journal of computer science and technology 27.3 (2012): 599-610.

Zou, D., Xie, H., Li, Q., Wang, F. L., & Chen, W. (2014). The Load-Based Learner Profile for Incidental Word Learning Task Generation. *In Advances in Web-Based Learning–ICWL* 2014 (pp. 190-200). Springer International Publishing.