A WordNet-Based Near-Synonyms and Similar-Looking Word Learning System

Koun-Tem Sun¹, Yueh-Min Huang² and Ming-Chi Liu²

¹Department of Information and Learning Technology, National University of Tainan, Taiwan // ²Department of Engineering Science, National Cheng Kung University, Taiwan // ktsun@mail.nutn.edu.tw // huang@mail.ncku.edu.tw // n9895113@mail.ncku.edu.tw

*corresponding author

ABSTRACT

Near-Synonyms and Similar-Looking (NSSL) words can create confusion for English as Foreign Language Learners as a result of a type of lexical error that often occurs when they confuse similar-looking words that are near synonyms to have the same meaning. Particularly, this may occur if the similar-looking words have the same translated meaning. This study proposes a method to find these NSSL words and designed three experiments to investigate whether NSSL matching exercises could increase Chinese EFL learners' awareness of NSSL words. Three primary findings arose from the study. First, a performance evaluation of the experiment showed good results and determined that the method extracted suitable NSSL words whose meaning EFL learners may confuse. Secondly, the analysis results of the evaluation of Computer Assisted Language Learning (CALL) software showed that this system is practical for language learning, but lacks authenticity. Thirdly, a total of ninety-two Chinese students participated in this study and the findings indicated that students increased awareness of NSSL words and improved in ability of NSSL word distinction while still maintaining the knowledge one month after they had completed the matching exercises. Additionally, students' feedback expressed that they had benefited from discovery learning and that they thought it was not difficult to discover the differences among NSSL words. Further research might extend the method proposed in this study to distracter choice of automatic question generation.

Keywords
Vocabulary learning, Discovery learning, WordNet, Near-synonyms, Similar-looking words

Introduction

Vocabulary knowledge is essential for reading texts in a foreign language (Chen & Hsu, 2008). Language teachers generally train students to make use of two important skills to understand the meaning of unknown words: (a) the skill of guessing (i.e., the ability to use contextual and structural clues to obtain the correct meaning of the words); (b) the ability to efficiently use the dictionary (i.e., to quickly refer to the page that contains the target word and to read the interpretation of the word).

Many vocabulary learning systems, such as the electronic dictionary and corpus-based software, have been developed for learning English vocabulary. With the electronic dictionary, learners can quickly look up an English word. The electronic dictionary not only reduces the learner’s querying time but also provides the same information as non-electronic dictionaries (Nagata, 1999). Beyond the electronic dictionary, corpora can be consulted as they provide more linguistic information than the electronic dictionary, such as how frequently words occur, which words tend to co-occur, and how the language is structured. The research of Cobb (1997) and Lee and Liou (2003) showed that a corpus-based approach can be used to scaffold the learning of students with low vocabulary skills and improve the results of vocabulary acquisition.

However, a word is described as a set of features. It is necessary to acquire the various features of a word in order to master that word completely. This includes features of its spoken and written forms, meanings, grammatical behavior, associations, collocations, frequency, and register (Richards, 1976). Many learners use the electronic dictionary only for obtaining one word with the simple translated gloss of the native language (see Figure 1). The inquirer only acquires the superficial meaning and is very apt to forget it (Martin, 1984). Heavy reliance on electronic dictionaries will negatively influence English reading fluency (Yeung, Jin, & Sweller, 1998). Another digital source, a corpus, can be consulted with concordance software. Most corpus-based systems focus on lexical choice, sorting the results based on their frequency and diction within the database queried. Only a few studies have attempted to use lexical structures, such as those pertaining to semantic relations for learning. Shimodaira et al. (2006) pointed out that some students queried words, whose meaning they already knew, to find the lexical relations...
of those words with other words. Nevertheless, these systems neglect the fact that a learner’s native language is one of the most important factors that may negatively influence the learning of foreign language vocabulary.

English has an alphabetic writing system that emphasizes the knowledge of sound-to-spelling correspondence. The Chinese characters have evolved from their earliest form of hieroglyphs that emphasize word-specific knowledge in pronunciation and spelling (Baron & Strawson, 1976). The meaning of a Chinese character may directly be inferred by part of the word. The results of neorolinguistic research revealed that Chinese native speakers perceive reading Chinese characters in both hemispheres of the brain. However, English native speakers perceive reading English words only in the left hemisphere of the brain. This research proves that there is a difference in the cognition of English words versus Chinese characters, and that this has an influence on learning (Cheng & Yang, 1989; Tzeng, Hung, Cotton, & Wang, 1979). On the other hand, because of the environmental limitations of their L1, Chinese students are relatively weak in sound-to-spelling competency in English. Therefore, for the Chinese EFL student, competency in sight-word recognition is more important. These students prefer organizing similar-looking words into groups to aid recall. Learning groups of similar looking words together can form handy chunks in memory to aid recall (Ellis, 1996). Nevertheless, with regard to some near-synonyms and similar-looking (NSSL) words, Chinese learners are often confused and find it difficult to learn both at the same time.

Two NSSL words can be defined as two similar-looking words that have a similar meaning. However, they are not synonyms, and thus cannot be substituted one for the other. For example, the two words, “transform” and “transfer,” are NSSL words. First, they have the same initial, “transf.,” second, the dictionary shows that they share the meaning of ‘change’ (as shown in Figure 1). But a more precise look reveals that they have different concepts and are only near-synonyms. The word “transform” is defined as, “to change somebody or something completely,” and the word “transfer” is defined as, “to move from one place to another.” (see Figure 2). Confusion of near-synonyms and similar-looking (NSSL) words is a type of semantic lexical error that Chinese EFL students typically make. The Chinese learner will likely confuse their meaning as a result of learning groups of NSSL words together.

Therefore, the focus of this research is to develop a semantically-related NSSL word learning system, and then to apply discovery-learning theory to help Chinese students practice English vocabulary.
WordNet semantic lexicon and discovery learning

WordNet and semantic relationships

WordNet is an English lexical database developed under the direction of Miller (Fellbaum, 1998). Its main content is made up of two major parts; one part contains sets of synonyms called “synsets,” and the other contains the semantic relations between these synonym sets. These two parts are explained as follows:

- **Synsets:** WordNet organizes English words into sets of synonyms called synsets, the minimal unit of the WordNet lexicon. When a word participates in several synsets, the relationship among those words is called a polysemy. WordNet quantifies this by frequency score. Synsets are displayed in the order of frequency.

- **Semantic relations:** Most synsets are connected to other synsets via semantic relations. These semantic relations are organized into WordNets, such as synonymy/antonymy, hypernymy/hyponymy, and meronymy/holonymy. For example, Figure 3 shows that the word “brother” is an antonym of the word “sister,” the phrase “organic substance” is a hyponym of the word “substance,” and the two words “arm” and “leg” are meronyms of the word “body.”

![Figure 3. Network representations of three semantic relations in WordNet (Miller, 1990)](image)

The WordNet lexicon offers a good semantic structure for computing the semantic similarity between words. The semantic structure of WordNet can be regarded as a tree graph. The nodes of the tree graph are synsets and the edges of the tree graph are semantic relations. Our research measured the semantic relations by calculating the lowest super-ordinate depth (Wu & Palmer, 1994). The concept behind this method is, as the depth of the lowest super-ordinate of the two concepts becomes deeper, the similarity of the concepts increases.

![Figure 4. An example of a similarity tree (Resnik, 1999)](image)

Besides Wu and Palmer's edge-counting approach, there is another type of semantic similarity on the WordNet hierarchy called the information-based approach (Resnik, 1999). Resnik's approach is a method of measuring semantic similarity between two concepts in an IS-A taxonomy; that is the degree to which they share information in common. From the perspective of teachers’ practical teaching experience, Wu and Palmer’s semantic similarity was chosen because calculating the lowest super-ordinate depth allows for teacher intuition in identifying the difficulty in distinguishing two similar concepts. For example, in Figure 4, if a student fails to distinguish between "dime" and "credit card,” it may be because he/she did not sufficiently learn the more abstract concept "medium of exchange." In this case, teachers would suggest the student study the less difficult concepts "money" and "credit" more thoroughly to learn "medium of exchange" before attempting "dime" and "credit card." However, the information-based
approach, with less use of the structure of the taxonomy, may increase the teachers’ difficulty in distinguishing students’ possible misconceptions.

Combining discovery learning theory with the NSSL word learning system

Discovery learning theory was proposed by Bruner in the 1960s (Bruner, 1971). This theory emphasizes that learning is not the passive receiving of knowledge, but that it is active discovery connecting various kinds of information. Novel knowledge is acquired by inductive reasoning. By linking novel knowledge with existing knowledge, learners can distinguish the difference between concepts. Through problem-solving activities, learners can memorize longer items, have a deeper understanding, and experience a smoother transfer of learning.

Our research developed an English NSSL word learning system based on the discovery learning theory. The operating procedure of the proposed system is shown below:
1. The user inputs a query word and a corresponding sentence that includes this word.
2. The system then shows a group of sentences containing NSSL words and the definitions of these words.
3. The user matches the sentences with the definitions by intuitive thinking.
4. The system provides feedback from the results of step 3.

In step 2, the proposed system filters out a group of NSSL words via our proposed method thereby avoiding learners having to rely only on blind discovery. In addition, the system provides matching items that are a group of sentences and the corresponding definitions. The matching exercise gives learners the opportunity to distinguish between some confusing concepts. This approach allows learners to actively solve problems via inductive reasoning and not simply be passive recipients. Moreover, this approach causes learners to link novel knowledge with existing knowledge through deeper thought processes. Consequently, the learning process in this system can assist learners not only in improving their retention and understanding of words, but also in guessing the meaning of the word in a similar context, thus, resulting in a better transfer of learning.

Method for finding near-synonyms and similar-looking words

By applying the proposed learning system, the user can input a query word to find its NSSL words. The proposed method is shown in Figure 5.

**Figure 5.** The process of the NSSL word learning system

Part of speech tagging (POS tagging)

This step parses all words of the input sentence. A POS tagger will mark up each word to a particular POS (such as verb, noun, adjective, or adverb). Our procedure uses Brill’s tagger (Brill, 1992); first it uses statistical techniques to extract information from training corpus and then uses a program to automatically learn rules. Brill’s tagger has an
accuracy of up to 97.2% when trained in contextual rules of 600,000 words in the Penn Treebank tagged Wall Street Journal Corpus with a 150,000 word test set (Brill, 1995).

Stop word check

Some words often appear in a sentence, but they are extremely common and semantically non-selective in the text. These words, called “stop words” (Manning, Raghavan, & Schütze, 2007), include such words as subject pronouns (I, you, he), prepositions (in, to, at), and conjunctions (and, but, while). WordNet does not include stop words in its database, so these words are shown to the user as stop words and then the system returns to the input query step.

Stem

The stem of a word usually inflects its ending to a derived word according to its grammatical function in a sentence. For example, word groups such as "transform," "transformed," and "transformation" are conflated into a single stem. However, words with a common stem will usually have similar meanings. In order to make the search faster and more precise, our procedure uses the Porter Stemming algorithm (Porter, 1980) to reduce the derived word to its stem.

Word sense disambiguation (WSD)

A word may have more than one sense, which leads to ambiguity. Word sense disambiguation is the process of finding out the most appropriate sense of a word for its use in a given sentence. Lesk’s algorithm (Lesk, 1986) is often used in knowledge-based disambiguation. This algorithm uses WordNet as its comparison dictionary, which provides the possible senses, definitions, and examples of a word. Lesk's algorithm is processed in three major steps: first, utilizing WordNet to search for polysemy in the query sentence and retrieve every sense of polysemy. Second, compare examples and definitions of each sense to the query sentence and count the number of words that are the same between the two groups. Third, the polysemic word is assigned to the sense whose examples and definitions have the largest number of common sense meanings in the query sentence. Seventy percent accuracy can be achieved with Lesk’s algorithm when using a typical learner’s dictionary. In addition to the automatic WSD, the frequency score provided by WordNet, a count indicating how often a word appears in a specific sense of all its polysemic senses, assists users in choosing the correct sense.

Finding near-synonyms

There are two steps in this procedure:
1. Use breadth-first searching (BFS) to find neighboring words around the query word. Theses neighboring words are related to each other through WordNet hypernymy-hyponymy relation. The maximal search depth is set to four in order to avoid an enormous search space.
2. Use Wu and Palmer’s semantic similarity equation (Wu & Palmer, 1994) [Eq. (1)] to measure the similarity between the query word and the neighboring words found in step 1. One hundred words with the greatest values of \( \text{sim}_{wp}(c_1, c_2) \) [see Eq. (1)] are selected as candidate words for the next procedure.

\[
\text{sim}_{wp}(c_1, c_2) = \frac{2 \times \text{depth}(\text{Iso}(c_1, c_2))}{\text{len}(c_1, \text{Iso}(c_1, c_2)) + \text{len}(c_2, \text{Iso}(c_1, c_2)) + 2 \times \text{depth}(\text{Iso}(c_1, c_2))}, \tag{1}
\]

where \( c_1 \) is a concept or a word,
\( \text{depth}(c_i) \) is the shortest distance from the root to node \( c_i \),
\( \text{len}(c_1, c_2) \) is the path length from node \( c_1 \) to node \( c_2 \),
and \( \text{Iso}(c_1, c_2) \) is the lowest super-ordinate of node \( c_1 \) and node \( c_2 \).

For example, to determine the semantic similarity between “transform” and “transfer,” this procedure first finds the lowest super-ordinate (LSO) of “transform” and “transfer” in the hypernymy-hyponymy semantic tree graph (see
Figure 6), and “change” is the LSO with $depth(change) = 1$. It can then obtain the $len(transform, change) = 1$, and $len(transfer, change) = 2$. From Eq.(1), the semantic similarity of “transform” and “transfer” is equal to 0.4.

![Diagram of hypernymy-hyponymy semantic structure of “transform” and “transfer”](image)

**Finding similar-looking words**

The one hundred near-synonyms obtained through the previous procedure are used as the input of this procedure. The two most similar-looking words are found by computing the Levenshtein distance (also called the edit distance) (Levenshtein, 1966). The edit distance between two strings is defined as the minimum number of operations needed to transform a source string into a target string, where an operation may be an insertion, deletion, or substitution of a single character. The smaller the edit distance, the more similar two words are in appearance. Only two words with the smallest edit distance and higher frequency score are selected, considering that the clarification of context sentences retrieved from WordNet has the limit of length and quantity. Too many words may increase the difficulty of matching exercises in the following step and may cause difficulty for the learners in guessing the meaning of the word.

**Match examples and definitions of NSSL**

After obtaining a pair of NSSL words, our procedure retrieves definitions and examples of these two words from WordNet and displays them as a matching exercise for the learner. The learner needs to intuitively identify the definition of the word in each example. Based on the outcome of the matching exercise, the proposed system gives the correct match as feedback to improve the learner’s vocabulary knowledge.

**Implementation**

Figure 7 shows the graphical user interface of the NSSL word learning system. First, a learner inputs a query sentence [step (1)]. In step (2), the system parses the query sentence into separate words. The learner chooses a query word and enters step (3). In step (3), the learner matches examples with definitions. When the learner has finished the matching exercise, the system reveals the correct answers to the learner.

**Experiments and evaluations**

In order to prove the performance of this system, to evaluate this CALL software from teachers’ perspectives, and to test the system’s practicality by students’ experience, this study conducted three kinds of experiments. The first experiment was the evaluation of NSSL word mining performance in terms of precision and recall. The second experiment applied Chapelle’s (2001) six principles to guide CALL evaluation. The third experiment was the evaluation of students’ learning performance with achievement tests and their learning attitude with a questionnaire.
Performance evaluation

The English reference word list used to evaluate the proposed NSSL word-finding method consists of 6,480 words collected by the College Entrance Examination Center, Taiwan (http://www.ceed.edu.tw/Research/paper_doc/ce37/ce37.htm). After extracting the verbs, removing those words found in WordNet but not in the English reference word list, and removing low frequency words, ninety words were chosen for this study. As mentioned in the section, “Method for finding near-synonyms and similar-looking words,” finding NSSL words should consider two important parameters, semantic similarity and edit distance. If each parameter is either too high or too low, the located words will not fit the definition of NSSL words. In this experiment, a few sample words were used to test with different settings and the resulting words were discussed with an English teaching expert. The results show that the semantic similarity (ss) is set to a range (0.6 ≤ ss < 0.7) and edit distance (ed) is set to a range (0.7 ≤ ed ≤ 1), which could find the most suitable NSSL words whose meaning EFL learners may confuse.
The evaluation of NSSL word-finding performance is expressed in terms of precision and recall, as shown below, where \( W \) is the two word pair that students will likely confuse the meaning due to being near-synonyms and similar-looking:

\[
\text{Precision} = \frac{\# \text{ of sample correctly extracted as } W}{\# \text{ of all samples output being } W} \times 100\%
\]

\[
\text{Recall} = \frac{\# \text{ of sample correctly extracted as } W}{\# \text{ of all samples that hold the target words } W} \times 100\%
\]

The evaluation procedure was as follows. First, three college English teachers were asked individually to find the corresponding NSSL words from ninety words with the help of the NSSL word system. Next, precision and recall results were summarized via the teachers' majority voting. Finally, the resulting NSSL words from the system were collected by means of the NSSL word-finding method from the same ninety words and were compared with the teachers' results.

The experimental results show that the precision of this system was about 83.12% and recall was about 77.11%. Therefore, the NSSL words found through this approach have a high probability of causing confusion. Additionally, this approach found most of the NSSL words in this dataset. In order to confirm the mutuality within each group of NSSL words, the role of the NSSL word was interchanged in each group of NSSL words. For example, if the querying word is A, and the result word is B, the role of B is changed as an input querying to test whether the system finds A or not. The results obtained a 90.63% accuracy rate for mutuality, revealing that this NSSL finding approach is highly consistent.

Observed factors that decreased the precision of the NSSL finding method are as follows:

- Some pairs of words, such as even/level that are considered near-synonyms, could be accurately substituted one for the other in the general context. For example, in the sentence, “he evens/levels the ground,” the pair of words can substitute one for the other. This system also found variants of some words, such as prize/prise, and gray/grey. Thus, these types of interchangeable words could not be recognized as easily confused NSSL words.

- There are some NSSL words that differ due to their prefixes, such as ascend/descend, and number/outnumber. From the perspective of the system's design, these types of words match the definition of NSSL words, having both characteristics of near-synonyms and similar-looking. However, from the perspective of teachers' practical teaching experience, some teachers argued that students could learn to recognize affixes first to assist them in differentiating between the meanings of the different prefixed NSSL words. Their divergence of opinions caused the precision to decrease.

Out of these two factors, the first is more often responsible for the decrease of the precision than the second one.

The main factor causing a decrease in recall were neighboring words that were in the selected range of Wu and Palmer’s semantic similarity but did not satisfy the selected range of the edit distance. This study could not set a range of semantic similarity and edit distance within a time to extract all of the target words.

\[
\text{Table 1. The evaluation of the NSSL word learning system}
\]

<table>
<thead>
<tr>
<th>Qualities</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language learning potential</td>
<td>Good</td>
</tr>
<tr>
<td>Meaning focus</td>
<td>Good</td>
</tr>
<tr>
<td>Learner fit</td>
<td>Good</td>
</tr>
<tr>
<td>Authenticity</td>
<td>Not authentic</td>
</tr>
<tr>
<td>Positive impact</td>
<td>Positive</td>
</tr>
<tr>
<td>Practicality</td>
<td>Good</td>
</tr>
</tbody>
</table>

**CALL software evaluation**

The same three college English teachers that participated in the performance evaluation applied Chappelle's six criteria to evaluate the NSSL word learning system. The criteria were intended to evaluate not only the software but
also the task-appropriateness. In this evaluation, a questionnaire based upon a six-point rating-scale was developed to rate each of Chappelle’s six criteria to determine their appropriateness. Each rating-scale item has some guided sub-questions proposed by Chappelle (2001) to assist teachers’ judgment in each criterion. Table 1 shows the results of their combined ratings. How the NSSL word learning system meets these criteria is discussed below.

- **Language learning potential**, the degree of opportunity present for beneficial focus on form: the task provided by the NSSL word learning system is the matching test where the learner can first analyze and understand the difference in the forms and the meanings of the particular NSSL words in focus, and then use the context of example sentences to practice the word usage. The system then gives the correct match as feedback on the matching exercise. A learner can focus both on the similar looking forms of given NSSL words and on how to distinguish between their meanings.

- **Meaning focus**, the extent to which the learners’ attention is directed toward the meaning of the language: as this paper has already mentioned in the “Language learning potential” criteria, since learners need to comprehend the meanings of the definitions and examples of NSSL words to do the matching exercises, there is a focus on meaning.

- **Learner fit**, the amount of opportunity for engagement with the language under appropriate conditions given learner characteristics: this activity was tailored to the learners’ vocabulary level in that the proposed system could be set to limit the located NSSL words to the English reference word list collected by the College Entrance Examination Center, Taiwan. This word list classifies 6,480 words into six levels selected on the basis of word frequency and the level of English or Chinese cultural background required to understand the word.

- **Authenticity**, the degree of correspondence between the learning activity and target language activities of interest to learners outside of the classroom: The task planned by teachers focused on some systemic NSSL words for students’ vocabulary growth. This task was not intended to be like tasks that learners would engage in outside the classroom unless learners choose to do so specifically for individualized vocabulary study. Therefore, the learners may not see this task of finding NSSL words as an authentic match with their future language use.

- **Positive impact**, the positive effects of the CALL activity on those who participate: Instructors and learners tended to have positive teaching and learning experiences with the NSSL word-learning system because it mined many relevant, potentially confusing words and attached matching exercises for distinguishing between these words, thereby enriching the range and flexibility of classroom materials.

- **Practicality**, the adequacy of the resources to support the use of the CALL activity: this system is developed on the web, so it does not need to be installed. The on-screen instructions and help information is clear and readily accessible. Most of the students had no difficulty in using the system.

**Students’ learning evaluation**

The hypothesis predicts that learners will perform well in differentiating between the different meanings of the NSSL words when they use the proposed NSSL word learning system. This study also asks the students for feedback on the online materials.

A one-group pre-posttest design was adopted to address the issue under investigation. Twenty graduate students and seventy-two undergraduate students who participated the entire time period of the experiment (over three months) were selected for this study. Most of the students received formal English instruction for six years prior to their junior and senior years of high school. The participants took either “Advanced Digital Learning” or “Applications of Internet” as an elective course. Both courses were conducted in computer laboratories and reviewed a topic related to computer-assisted learning. Before conducting this experiment, the lectures introduced the topic of computer-assisted learning, and then this experiment was carried out as an example of the introduced topic. Subjects were told that their experiment outcome will not impact their final grades; therefore, participation in this study was voluntary.

Two types of instruments, a test and a questionnaire, were used to collect data for the educational research hypothesis. The pretest contained 15 groups of NSSL target words (a total of 30 words) and 10 distracters. The 20 sentence-completion test items that made up the pretest, posttest, and delayed posttest were identical but sequenced differently in each of the tests. The questionnaire had 10 items for examining students’ perception of using the system after the experiment. Each item had a five-point Likert-type scale.

The procedure is as follows. First, all participants received a training course for fifteen minutes on how to use the proposed NSSL word learning system, and then they were invited to take the pretest in the first week. Next, in the
five-week treatment stage, teachers taught the definitions and gave examples of the target words. Students went on to query the target word by using the NSSL word learning system, and then they completed the matching exercises of the NSSL word learning system that applied discovery learning strategy in the exercises. They were given 5 minutes of in-class time and had to complete the rest after class each week. In total, 30 words, that is to say 15 pairs of NSSL words, formed the target word list that was equally distributed over the five units, one unit per week. Last, in the posttest stage, students took the immediate posttest and filled out the evaluation questionnaire. Then, four weeks later, students took the delayed posttest. The results of the tests and student feedback on the system are presented and discussed below.

Learners’ Performance and Retention as Measured in Controlled Tests

Comparisons were made to see if there were significant differences between (1) the scores of the pretest and the immediate posttest (see Table 2) and (2) the scores of the immediate posttest and the delayed posttest. The posttest scores (in which each item was assigned 5 points, 75 points in total for each test) were significantly higher than those of the pretest ($p=0.000<0.01$), and no significant difference was found in comparing the immediate posttest and the delayed posttest ($p>0.01$). Hence, the positive results indicate that students’ knowledge of NSSL words had increased significantly in the controlled tests. Additionally, students generally did not show much regression in the delayed posttest, as indicated by the comparison of both posttest scores. The results confirmed our hypothesis that the NSSL word-learning system did enhance students’ learning of NSSL words, and the effects were maintained over four weeks when learning was again measured by the same test items.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediate Posttest – Pretest</td>
<td>18.12</td>
<td>6.91</td>
<td>.000</td>
</tr>
<tr>
<td>Delayed Posttest – immediate Posttest</td>
<td>-.07</td>
<td>7.86</td>
<td>.937</td>
</tr>
</tbody>
</table>

Learners’ Perception of the NSSL Word-Learning System

The developed questionnaire was verified by conducting factor and reliability analysis. Conducting the principal components factor analysis, this study identified the 3 factors as follows:

- factor 1, explaining 35.97% of the variance, shows the usefulness of learning with the NSSL word-learning system;
- factor 2, explaining 12.79% of the variance, shows students’ perception of the design of the fifteen online learning practices;
- factor 3, explaining 10.44% of the variance, shows students’ overall attitude toward the matching exercise and their future use of the NSSL word-learning system.

The principal components explained 59.19% of the variance. The associated Cronbach alpha reliabilities, the principal components factor analysis, and students’ responses to the questions are shown in Table 3.

In the responses to the questionnaire, a great majority of students (95%) indicated that they liked NSSL word learning; a mere 5% held a neutral attitude. A total of 75% of the students strongly agreed (37%) or agreed (38%) that they will utilize this system to learn NSSL words in the future; a mere 25% of students held a neutral attitude. More than half of the students (62% and 61%) reported that they had no difficulty in making distinctions among NSSL words and in understanding example sentences. A total of 78% of the students (45% plus 33%) thought that the provided matching exercises helped them consider the differences among the NSSL words. These results concerning the, “perception of the design of the matching exercises” indicate that learning two NSSL words at a time in a matching exercise is sufficient for discovery learning and that the difficulty of the mined words by the proposed approach was adequate for the students’ English proficiency level. The overall results of the five questions pertaining to “usefulness of Learning” show that more than half of the students (57%, 68%, 56%, 61%, and 58%) responded that our system had helped them learn English NSSL words effectively. Furthermore, it seems that the NSSL word-learning system was more effective than using a typical dictionary.
Table 3. Results of students’ evaluation of the proposed system

<table>
<thead>
<tr>
<th>Questions</th>
<th>Percentage of Respondents % (n=92, α=0.78&gt;0.7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strongly Agree</td>
</tr>
<tr>
<td>Overall attribute &amp; future use</td>
<td></td>
</tr>
<tr>
<td>I liked trying to induce the exact meaning of NSSL words from the matching exercises provided by this system.</td>
<td>95</td>
</tr>
<tr>
<td>I will utilize this system to learn NSSL words in the future.</td>
<td>37</td>
</tr>
<tr>
<td>Perception of the design of the matching exercises</td>
<td></td>
</tr>
<tr>
<td>In discovery learning, I found it effortless to make distinctions between NSSL words.</td>
<td>9</td>
</tr>
<tr>
<td>In discovery learning, I felt that the example sentences were not difficult.</td>
<td>2</td>
</tr>
<tr>
<td>In discovery learning, I felt that doing the matching exercise helped me consider the subtle differences between the collocations of NSSL words.</td>
<td>45</td>
</tr>
<tr>
<td>Usefulness of learning</td>
<td></td>
</tr>
<tr>
<td>I feel that I understood the exact meaning of NSSL words better by doing the matching exercises in this system than I would have by looking up the word and reading the English or Chinese definition in a more typical dictionary.</td>
<td>33</td>
</tr>
<tr>
<td>I feel that, after submitting my answer, the feedback message to re-do the incorrect questions was helpful.</td>
<td>16</td>
</tr>
<tr>
<td>I feel that the layout of the online practice was easy to read.</td>
<td>22</td>
</tr>
<tr>
<td>I feel that the matching exercises helped me understand the word when reading it in context.</td>
<td>16</td>
</tr>
<tr>
<td>I feel that, after completing these exercises, I will be able to distinguish word meanings more precisely when reading.</td>
<td>14</td>
</tr>
</tbody>
</table>

Related works

Before presenting our conclusions and future work, one more point about related studies must be clarified. Because vocabulary test construction is time-consuming and expensive, some studies (Hoshino & Nakagawa, 2008; Jonathan, Gwen, & Maxine, 2005; Sumita, Sugaya, & Yamamoto, 2005) have tried to solve this problem by proposing a method of automatic question generation. There are two differences between their studies and ours in the applications of the research and in the methods of determining distracters. First, the two kinds of studies focused on different applications of the research. The automatic question generation studies tried to improve Computer-based testing (CBT), whereas our study tried to improve CALL. The two different applications of the research also resulted in two different experimental methods: the automatic question generation studies designed their experiment for the purpose of evaluating test validation (Tao, Wu, & Chang, 2008). Our study, on the other hand, designed the experiment for the purpose of evaluating the effectiveness of the learning. Second, in selecting similar words, most studies determined the distracters for their questions through machine-readable dictionaries (Hoshino & Nakagawa, 2008) such as an in-house thesaurus (Sumita et al., 2005) or WordNet (Jonathan et al., 2005). In the research using an in-house thesaurus, had a different thesaurus been used, the distracter candidates would most likely differ. On the other hand, in the research using WordNet, distracters were chosen directly from one-dimensional semantically related words (synonyms, antonyms, hypernyms, hyponyms, etc.) of the target words. In contrast, our study
determined the distracters for the matching exercises with the dynamic two-dimensional NSSL words (near-synonyms and similar-looking) through WordNet.

Conclusions and future work

While most vocabulary learning systems provide only one-dimensional semantically related words (synonyms, antonyms, hypernyms, hyponyms, etc.) of the target words, this study enhances the previous vocabulary learning systems by providing dynamic two-dimensional NSSL words (near-synonyms and similar-looking) through WordNet; then this research investigated whether NSSL matching exercises could increase EFL learners’ awareness of NSSL words. Our research questions first addressed the effectiveness of the proposed NSSL generating method and the proposed NSSL word learning system, and then addressed measurement of students’ learning gains in this system.

To address the first question, three college English teachers were invited to participate in this study in the aspects of performance evaluation and CALL software evaluation. After testing our proposed NSSL word mining method, their performance evaluation assigned good results. This method extracted suitable NSSL words whose meaning EFL learners are likely to confuse due to a misunderstanding with similar-looking words that are near synonyms and have the same meaning. Particularly if the similar-looking words have the same translated meaning the EFL learners would be confused about the exact meanings of the NSSL words. The results of the analysis of the CALL software evaluation showed that this system is good for practical language learning with the exception of the lack of authenticity. The lack of authenticity could be solved by teachers’ designing real-life material for students. However, CALL software analysis is always situation specific (Chapelle, 2001). Hence, the main purpose of applying CALL software analysis is providing a means for deciding whether to try the task in a language class or whether to attempt to improve the weak task.

To address the second question, twenty graduate and seventy-two undergraduate students were invited to participate in the study and offer their evaluation of the learning experience. Although the results showed that the students’ test scores had increased between pretest and posttest and still maintained the knowledge of NSSL words over one month, these findings do not say that the proposed system applied a discovery learning strategy that caused these learning effects. It would be better to say that the proposed NSSL word learning system facilitated EFL learners’ to increase the “awareness” of NSSL words. Further, they are willing to learn these kinds of NSSL words through the proposed match exercise or other preferred individual learning resources. Students’ overall feedback, which showed that they had a positive attitude toward the usefulness of the proposed system and were willing to use it in the future, may also explain the benefit of the NSSL word learning system.

A pedagogical implication can be drawn for EFL teachers from this study. Unlike homophones, NSSL words do not seem to get any special treatment in the vocabulary curriculum, perhaps because of a lack of awareness of the particular features of similarity that lead to the confusion between them. The students usually use a dictionary or the skill of educated guessing to find the meaning of unfamiliar words. When, on the other hand, the words look similar, or are near-synonyms to more familiar words, they may not waste their time on checking them. The importance of this erroneous strategy should not be overestimated in the case of words that seem familiar to the students. English teachers of Chinese students should spend more of their teaching time on distinguishing the exact meaning between these NSSL words.

It is important to note that this method of investigation is not thoroughly flawless. First, the study design does not record the differences in student performance with respect to the time he or she has spent with the learning system. It is not known how many learning effects were caused by the proposed system. Secondly, considering the clarification of context sentences retrieved from WordNet has a limit of length and quantity the matching exercises only provide two matching items at a time. Any extra resources, such as corpus, could be used to enrich the context sentences of WordNet. Finally, a larger number and variety of participants is needed to make further generalizations of results. Further research, that includes these and other issues, needs to be conducted to draw more rigorous claims.

There are two potential issues for future study. One is the possibility of extending and applying the methods used in this study to the distracter choice in automatic question generation. The other is applying other related ontology-
based similarity measures or string-based similarity measures such as SimPack (Bernstein, Kaufmann, Kiefer, & Bürki, 2005) to investigate the potential for more effective learning.

Acknowledgments

This work was supported in part by the National Science Council (NSC), Taiwan, ROC, under Grant NSC 97-2511-S-006-001-MY3 and NSC 98-2631-S-024-001.

References


