Developing and Evaluating a Computer-Assisted Near-Synonym Learning System

YU Liang-Chih HSU Kai-Hsiang
Department of Information Management, Yuan Ze University, Chung-Li, Taiwan, R.O.C.
lcyu@saturn.yzu.edu.tw, s986220@mail.yzu.edu.tw

ABSTRACT

Despite their similar meanings, near-synonyms may have different usages in different contexts. For second language learners, such differences are not easily grasped in practical use. In this paper, we develop a computer-assisted near-synonym learning system for Chinese English-as-a-Second-Language (ESL) learners using two automatic near-synonym choice techniques: pointwise mutual information (PMI) and *n*-grams. The two techniques can provide useful contextual information for learners, making it easier for them to understand different usages of various English near-synonyms in a range of contexts. The system is evaluated using a vocabulary test with near-synonyms as candidate choices. Participants are required to select the best near-synonym for each question both with and without use of the system. Experimental results show that both techniques can improve participants' ability to discriminate among near-synonyms. In addition, participants are found to prefer to use the PMI in the test, despite *n*-grams providing more precise information.

KEYWORDS : Near-synonym choice, computer-assisted language learning, lexical semantics

1 Introduction

Near-synonym sets represent groups of words with similar meanings, which can be derived from existing lexical ontologies such as WordNet (Fellbaum, 1998), EuroWordNet (Rodríguez et al., 1998), and Chinese WordNet (Huang et al., 2008). These are useful knowledge resources for many applications such as information retrieval (IR) (Moldovan and Mihalcea, 2000; Navigli and Velardi, 2003; Shlrl and Revle, 2006; Bhogal et al., 2007) and computer-assisted language learning (CALL) (Cheng, 2004; Inkpen, 2007; Ouyang et al., 2009; Wu et al., 2010). For instance, in CALL, near-synonyms can be used to automatically suggest alternatives to avoid repeating the same word in a text when suitable alternatives are available in its near-synonym set (Inkpen, 2007). Although the words in a near-synonym set have similar meanings, they are not necessarily interchangeable in practical use due to their specific usage and collocational constraints (Wible et al., 2003; Futagia et al., 2008). Consider the following examples.

(1) {strong, powerful} coffee (Pearce, 2001)
(2) ghastly {error, mistake} (Inkpen, 2007)

Examples (1) and (2) both present an example of collocational constraints for the given contexts. For instance, in (1), the word *strong* is more suitable than *powerful* in the context of "coffee", since "powerful coffee" is an anti-collocation. These examples indicate that near-synonyms may have different usages in different contexts, and such differences are not easily captured by second language learners. Therefore, this study develops a computer-assisted near-synonym learning system to assist Chinese English-as-a-Second-Language (ESL) learners to better understand different usages of various English near-synonyms.

To this end, this study exploits automatic near-synonym choice techniques (Edmonds, 1997; Inkpen, 2007; Gardiner and Dras, 2007, Islam and Inkpen, 2010; Wang and Hirst, 2010; Yu et al., 2010a; 2010b; 2011) to verify whether near-synonyms match the given contexts. Figure 1 shows an example of near-synonym choice. Given a near-synonym set and a sentence containing one of the near-synonyms, the near-synonym is first removed from the sentence to form a lexical gap. The goal is to predict an answer (i.e., best near-synonym) to fill the gap from the near-synonym set according to the given context. The *pointwise mutual information (PMI)* (Inkpen, 2007; Gardiner and Dras, 2007), and *n-gram* based methods (Islam and Inkpen, 2010; Yu et al., 2010b) are the two major approaches to near-synonym choice. PMI is used to measure the strength of co-occurrence between a near-synonym and individual words appearing in its context, while n-grams can capture contiguous word associations in the given context. Both techniques can provide useful contextual information for the near-synonyms. This study uses both techniques to implement a system with which learners can practice discriminating among near-synonyms.

Sentence: This will make the **_____** message easier to interpret. (Original word: error)
Near-synonym set: {error, mistake, oversight}

FIGURE 1 – Example of near-synonym choice.

2 System Description

2.1 Main Components

1) PMI: The pointwise mutual information (Church and Hanks, 1991) used here measures the co-occurrence strength between a near-synonym and the words in its context. Let w_i be a word in the context of a near-synonym NS_j. The PMI score between w_i and NS_j is calculated as

$$PMI(w_i, NS_j) = \log_2 \frac{P(w_i, NS_j)}{P(w_i)P(NS_j)}, \tag{1}$$

where $P(w_i, NS_j) = C(w_i, NS_j)/N$ denotes the probability that w_i and NS_j co-occur; $C(w_i, NS_j)$ is the number of times w_i and NS_j co-occur in the corpus, and N is the total number of words in the corpus. Similarly, $P(w_i) = C(w_i)/N$, where $C(w_i)$ is the number of times w_i occurs, and $P(NS_j) = C(NS_j)/N$, where $C(NS_j)$ is the number of times NS_j occurs. All frequency counts are retrieved from the Web 1T 5-gram corpus. Therefore, (1) can be re-written as

$$PMI(w_i, NS_j) = \log_2 \frac{C(w_i, NS_j) \cdot N}{C(w_i)C(NS_j)}. \tag{2}$$

The PMI score is then normalized as a proportion of w_i occurring in the context of all near-synonyms in the same set, as shown in Eq. (3).

$$\widetilde{PMI}(w_i, NS_j) = \frac{PMI(w_i, NS_j)}{\sum_{j=1}^{K} PMI(w_i, NS_j)}, \tag{3}$$

where $\widetilde{PMI}(w_i, NS_j)$ denotes the normalized PMI score, and K is the number of near-synonyms in a near-synonym set.

2) N-gram: This component retrieves the frequencies of n (2~5) contiguous words occurring in the contexts from the Web 1T 5-gram corpus.

2.2 System Implementation

Based on the contextual information provided by the PMI and N-gram, the system implements two functions: contextual statistics and near-synonym choice, both of which interact with learners. The system can be accessed at http://nlptm.mis.yzu.edu.tw/NSLearning.

1) Contextual statistics: This function provides the contextual information retrieved by PMI and N-gram. This prototype system features a total of 21 near-synonyms grouped into seven near-synonym sets, as shown in Table 1. Figure 2 shows a screenshot of the interface for contextual information lookup. For both PMI and N-gram, only the 100 top-ranked items are presented.

2) Near-synonym choice: This function assists learners in determining suitable near-synonyms when they are not familiar with the various usages of the near-synonyms in a given context. Learners can specify a near-synonym set and then input a sentence with "*" to represent any near-synonym in the set. The system will replace "*" with each near-synonym, and then retrieve the contextual information around "*" using PMI and N-gram, as shown in Fig. 3. For PMI, at most five context words (window size) before and after "*" are included to compute the normalized PMI scores for each near-synonym. In addition, the sum of all PMI scores for each near-synonym is also presented to facilitate learner decisions. For N-gram, the frequencies of the n-grams (2~5) containing each near-synonym are retrieved.

No.	Near-Synonym sets	No.	Near-Synonym sets
1	difficult, hard, tough	2	error, mistake, oversight
3	job, task, duty	4	responsibility, burden, obligation, commitment
5	material, stuff, substance	6	give, provide, offer
7	settle, resolve		

TABLE 1 – Near-synonym sets.

Near-synonym set job task duty

job			task			duty		
Context	PMI_score	Frequency	Context	PMI_score	Frequency	Context	PMI_score	Frequency
teen	1	1,816,316	trivial	1	78,286	cycle	1	342,491
seekers	1	1,479,452	committees	1	75,321	breach	1	336,947
listings	1	1,473,629	pane	1	52,660	fiduciary	1	325,983
opportunities	1	1,416,347	privileged	1	49,161	tour	1	240,835
openings	1	1,071,984	force	0.99	2,874,435	stamp	1	172,109

Near-synonym set job task duty N-gram 4 Submit

job		task		duty	
to do the job	438,769	the task at hand	172,859	have a duty to	202,441
did a great job	425,841	with the task of	167,026	is the duty of	191,024
a good job of	412,589	not an easy task	145,907	be the duty of	178,800
do a better job	357,618	up to the task	143,106	shall be the duty	161,951
link to save job	346,000	of the task force	122,120	the line of duty	160,011

FIGURE 2 – Screenshot of contextual statistics.

Near-Synonym set material stuff substance

It was found that the * of the matter and not only mere theory was to be regarded submit

Window size 3

PMI	material	stuff	substance
found	0.35	0.34	0.31
that	0.24	0.49	0.28
the	0.31	0.27	0.42
of	0.28	0.28	0.44
the	0.31	0.27	0.42
matter	0.15	0.08	0.77
	1.64	1.73	2.64

Bi-gram	the material	7,488,173	the stuff	2,581,457	the substance	1,319,583
	material of	817,776	stuff of	392,805	substance of	848,188
Tri-gram	that the material	237,330	that the stuff	25,305	that the substance	52,803
	the material of	129,580	the stuff of	254,931	the substance of	545,206
	material of the	179,643	stuff of the	31,130	substance of the	341,962
4-gram	found that the material	1,706	found that the stuff	211	found that the substance	623
	that the material of	3,082	that the stuff of	910	that the substance of	15,240
	the material of the	48,148	the stuff of the	9,307	the substance of the	242,832
	material of the matter	0	stuff of the matter	0	substance of the matter	6,205

FIGURE 3 – Screenshot of near-synonym choice.

3 Experimental Results

3.1 Experiment Setup

1) Question design: To evaluate the system, we designed a vocabulary test with near-synonyms as candidate choices. The vocabulary test consisted of 50 questions with a single correct answer for the 21 near-synonyms, where each near-synonym had at least two questions. The remaining eight randomly selected near-synonyms had three questions each. Each question was formed from a sentence selected from the British National Corpus (BNC). Figure 4 shows a sample question. For each question, the original word removed was held as the correct response.

Question:	He wanted to do a better _____ than his father had done with him. A. job B. task C. duty
Questionnaire 1:	How much did you depend on the system to answer the question? ☐ 1 (Not at all dependent) ☐ 2 ☐ 3 ☐ 4 ☐ 5 (Completely dependent)
Questionnaire 2:	Which method did you use in the test? ☐ PMI ☐ N-gram

FIGURE 4 – Sample question in the vocabulary test. The original word in the lexical gap is job.

2) Test procedure: In testing, participants were asked to propose an answer from the candidate choices, first in a pre-test without use of the system, and then in a post-test using the system. To obtain detailed results, participants were requested to provide two feedback items after completing each question, as shown in Figure 4. The first item is a 5-point scale measuring the degree to which the participant felt reliant on the system during the test, and reflects participants' confidence in answering questions. In the second item, participants were asked to indicate which method, PMI or n-grams (or both or none) provided the most useful contextual information.

3.2 Evaluation Results

A total of 30 non-native English speaking graduate students volunteered to participate in the test. Experimental results show that the participants scored an average of 44% correct on the pre-test. After using the system, this increased substantially to 70%. This finding indicates that the use of the system improved participants' ability to distinguish different usages of various near-synonyms. We performed a cross analysis of the two questionnaire items against the 1500 answered questions (i.e., 30 participants each answering 50 questions) in both the pre-test and post-test, with results shown in Table 2. The columns C_{pre}/C_{post}, $C_{pre}/\overline{C}_{post}$, $\overline{C}_{pre}/C_{post}$ and $\overline{C}_{pre}/\overline{C}_{post}$ represent four groups of questions partitioned by their answer correctness, where C_* and $\overline{C}_*$ respectively denote questions answered correctly and incorrectly in the pre-test or post-test. The rows labeled Without_system and With_system represent two groups of answered questions partitioned according to participants' ratings on the first questionnaire item, where Without_system represents ratings of 1 and 2, and With_system represents ratings of 3~5.

For Without_system, around 36% (536/1500) questions in the post-test were answered without use of the system due to high confidence on the part of participants. As shown in Fig. 5, around 59% (315/536) of these questions were answered correctly in both the pre-test and post-test, while only 28% (151/536) were answered incorrectly in both the pre-test and post-test, indicating that participants' confidence in their ability to answer certain questions correctly was not misplaced. The remaining 13% of questions provided inconsistent answers between the pre-test and post-test. For With_system, around 64% (964/1500) questions answered using the system in the post-test. Of these questions, around 46% (448/964) were answered incorrectly in the pre-test but were corrected in the post-test, indicating that participants had learned useful contextual information from the system. Around 25% (244/964) of questions answered correctly in the pre-

	C_{pre}/C_{post}	$C_{pre}/\overline{C}_{post}$	$\overline{C}_{pre}/C_{post}$	$\overline{C}_{pre}/\overline{C}_{post}$	Total	
Without_system	315	21	49	151	**536**	**1500**
With_system	244	78	448	194	**964**	
PMI	91	51	239	100	**481**	**824**
N-gram	93	19	177	54	**343**	

TABLE 2 – Cross analysis of questionnaire items against answered questions.

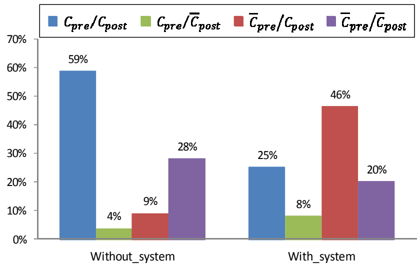

FIGURE 5 – Histograms of with and without system.

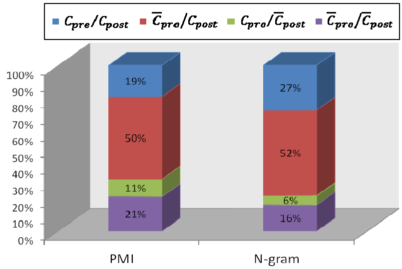

FIGURE 6 – Results of N-gram and PMI.

test were also answered correctly in the post-test because participants became more confident after double-checking their proposed answers with the system. Only 8% (78/964) of questions answered correctly in the pre-test were answered incorrectly in the post-test, and the remaining 20% of questions answered incorrectly in the pre-test were still incorrect in the post-test. A possible explanation is that the system does not always provide perfect results. In some cases, the system may provide ambiguous information, such as when the given context is too general. In such cases, participants may propose incorrect answers despite having used the system.

3.3 Comparison of PMI and N-gram

Table 2 shows that there were a total of 824 questions with feedback on the second questionnaire item, where 58% of questions were answered based on PMI, and 42% based on N-gram, indicating that participants had a preference for PMI in the test. But, in fact, previous studies have shown that the 5-gram language model has an accuracy of 69.9%, as opposed to 66.0% for PMI (Islam and Inkpen, 2010), thus N-gram provides more precise information. Evaluation results of 50 questions were consistent with this discrepancy, showing the respective accuracies of N-gram and PMI to be 68% and 64%. Figure 6 shows the comparative results of PMI and N-gram. The percentages of both C_{pre}/C_{post} and $\overline{C}_{pre}/C_{post}$ for N-gram were higher than those for PMI, and the percentages of both $C_{pre}/\overline{C}_{post}$ and $\overline{C}_{pre}/\overline{C}_{post}$ for N-gram were lower than those for PMI. Overall, N-gram use resulted in a correct/incorrect ratio of 79:21 in the post-test, as opposed to 69:31 for PMI, indicating that N-gram can assist participants in correctly answering more questions and producing fewer errors caused by ambiguous contextual information.

Conclusion

This study developed a computer-assisted near-synonym learning system using two automatic near-synonym choice techniques: PMI and N-gram, which can capture the respective individual and contiguous relationship between near-synonyms and their context words. Results show that both techniques can provide useful contextual information to improve participants' ability to discriminate among near-synonyms. While participants had a preference for PMI, *n*-grams can provide more precise information. Future work will be devoted to enhancing the system by including more near-synonym sets and incorporating other useful contextual information.

Acknowledgments

This work was supported by the National Science Council, Taiwan, ROC, under Grant No. NSC100-2632-S-155-001 and NSC99-2221-E-155-036-MY3.

References

Bhogal, J., Macfarlane, A., and Smith, P. (2007). A Review of Ontology based Query Expansion. *Information Processing and Management*, 43(4):866-886.

Cheng, C. C. (2004). Word-Focused Extensive Reading with Guidance. In *Proc. of the 13th International Symposium on English Teaching*, pages 24-32.

Church, K. and Hanks, P. (1990). Word Association Norms, Mutual Information and Lexicography. *Computational Linguistics*, 16(1):22-29.

Edmonds, P. (1997). Choosing the Word Most Typical in Context Using a Lexical Co-occurrence Network. In *Proc. of the 35th Annual Meeting of the Association for Computational Linguistics (ACL-97)*, pages 507-509.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Futagia, Y., Deanea, P., Chodorow, M., and Tetreault, J. (2008). A Computational Approach to Detecting Collocation Errors in the Writing of Non-native Speakers of English. *Computer Assisted Language Learning*, 21(4):353-367.

Gardiner, M. and Dras, M. (2007). Exploring Approaches to Discriminating among Near-Synonyms, In *Proc. of the Australasian Technology Workshop*, pages 31-39.

Huang, C. R., Hsieh, S. K., Hong, J. F., Chen, Y. Z., Su, I. L., Chen, Y. X., and Huang, S. W. (2008). Chinese Wordnet: Design, Implementation, and Application of an Infrastructure for Cross-lingual Knowledge Processing. In *Proc. of the 9th Chinese Lexical Semantics Workshop*.

Inkpen, D. (2007). A Statistical Model of Near-Synonym Choice. *ACM Trans. Speech and Language Processing*, 4(1):1-17.

Islam, A. and Inkpen, D. (2010). Near-Synonym Choice using a 5-gram Language Model. *Research in Computing Science: Special issue on Natural Language Processing and its Applications*, Alexander Gelbukh (ed.), 46:41-52.

Moldovan, D. and Mihalcea, R. (2000). Using Wordnet and Lexical Operators to Improve Internet Searches. *IEEE Internet Computing*, 4(1):34-43.

Navigli, R. and Velardi, P. (2003). An analysis of ontology-based query expansion strategies. In *Proc. of the Workshop on Adaptive Text Extraction and Mining (ATEM).*

Ouyang, S., Gao, H. H., and Koh, S. N. (2009). Developing a Computer-Facilitated Tool for Acquiring Near-Synonyms in Chinese and English. In *Proc. of the 8th International Conference on Computational Semantics (IWCS-09)*, pages 316-319.

Pearce, D. (2001). Synonymy in Collocation Extraction. In *Proc. of the Workshop on WordNet and Other Lexical Resources at NAACL-01*.

Rodríguez, H., Climent, S., Vossen, P., Bloksma, L., Peters, W., Alonge, A., Bertagna, F., and Roventint, A. (1998). The Top-Down Strategy for Building EeuroWordNet: Vocabulary Coverage, Base Concepts and Top Ontology, *Computers and the Humanities*, 32:117-159.

Shlrl, A. and Revle, C. (2006). Query expansion behavior within a thesaurus-enhanced search environment: A user-centered evaluation. *Journal of the American Society for Information Science and Technology*, 57(4):462-478.

Wang, T. and Hirst, G. (2010). Near-synonym Lexical Choice in Latent Semantic Space. In *Proc. of the 23rd International Conference on Computational Linguistics (Coling-10),* pages 1182-1190.

Wible, D., Kuo, C. H., Tsao, N. L., Liu, A., and Lin, H. L. (2003). Bootstrapping in a Language Learning Environment. *Journal of Computer Assisted Learning*, 19(1):90-102.

Wu, C. H., Liu, C. H., Matthew, H., and Yu, L. C. (2010). Sentence Correction Incorporating Relative Position and Parse Template Language Models. *IEEE Trans. Audio, Speech and Language Processing*, 18(6):1170-1181.

Yu, L. C., Chien, W. N., and Chen, S. T. (2011). A baseline system for Chinese near-synonym choice. In *Proc. of the 5th International Joint Conference on Natural Language Processing (IJCNLP-11),* pages 1366-1370.

Yu, L. C., Shih, H. M., Lai, Y. L., Yeh, J. F., and Wu, C. H. (2010a). Discriminative Training for Near-synonym Substitution. In n *Proc. of the 23rd International Conference on Computational Linguistics (Coling-10),* pages 1254-1262.

Yu, L. C. Wu, C. H. Chang, R. Y. Liu, C. H., and Hovy, E. H. (2010b). Annotation and Verification of Sense Pools in OntoNotes. *Information Processing and Management*, 46(4):436-447.