Evaluation on Second Language Collocational Congruency with Computational Semantic Similarity

Ching-Ying Lee **Chih-cheng Lin**
Department of English, National Taiwan Normal University
Department of Applied Foreign Languages, Kang Ning Junior College
collocation354@gmail.com

Abstract

Collocation learning is one of the important building blocks for the development of language competence. Remarkably, it is influenced by L1 and L2 congruency. The present study thus focused on the distinguishability of the computational similarity values between L2 collocates and L1 counterparts to establish the use of semantic similarity measure as a research instrument. The results showed that the inconsistency between human (subjective) and computational (objective) congruency classification of verb-noun collocations.

1 Introduction

Collocation learning is an important research area because it involves structural, semantic and cognitive variations in lexical components which underpin the foundation of language competence. The notion of collocational congruency distinguishes whether an L2 collocation is congruent or incongruent with L1 counterpart. Evaluation of collocational congruency is currently performed by human judgment. This subjective evaluation mostly depends on individual lexical knowledge and word meaning interpretation. Human judgment on meaning accordance lacks a clear criterion as to clear-cut L2 collocation in congruency. How much similar in word meaning can be considered as congruent collocation? How much different should be regarded as incongruent? This vagueness is not resolvable by human judgment and can only expect inconsistent evaluation.

The current study identified a research gap in the literature of L2 collocation where theoretical concepts of collocation congruency remain vague and lack explicit criteria for subjectively dichotomous congruency classification (Koya, 2005; Webb & Kagimoto, 2009; Yamashita & Jiang, 2010; Wolter & Gyllstad, 2011). This research proposed an objective and systematic method for congruency evaluation by exploiting computational measures of lexical semantic similarity. Based on literature review, it was found that *WordNet* (Miller, 1995) incorporated eight computational algorithms of semantic similarity measures and provided convenient online use. Of the eight algorithms, two (*Adapted Lesk* and *Gloss Vectors*) were selected in terms of their computational features and measuring stability. Three sets of word pairs with different semantic relations were composed and tested for lexical similarity values by the two measures.

The current study further applied the two similarity measures to the experimental set of collocation so as to objectively evaluate the properties of congruency. Semantic similarity between a collocate and an L2 transferred word with L1 word sense was quantified by the two computational similarity measures. Statistical and analytical comparisons were made, which led to further understanding of the potential advantage of exploiting semantic similarity for congruency evaluation.

2 Instruments and Data Collection

The research instruments included two semantic similarity measures and a set of collocation test. By the operational definition, a collocation is formed by a collocate and a base. Given a pair of equivalent L2 and L1 collocations, the subject of study is usually the semantic relation between the pair of cross-linguistic collocates. However, currently all semantic similarity measures were designed to operate on word pairs of the same

27th Pacific Asia Conference on Language, Information, and Computation pages 534－541

language. To evaluate the semantic relation between the pair of cross-linguistic collocates by semantic similarity measures, an L2 transferred word of the L1 collocate was used as a surrogate that embedded the word sense of the L1 collocate.

As a design feature, semantic similarity measures also allowed semantic similarity evaluation between word pairs in both contexts of all word senses or designated word senses. When operated in all word senses, semantic similarity measures computed all possible combinations of word senses and gave the highest value that reflected the most similar senses of the two words. Alternatively, when a particular sense of each word was selected, semantic similarity measures provided similarity values of the two designated word senses.

As a convenient and useful semantic search instrument, *WordNet Search-3.1* was employed to consult for word senses in glossary. The online system of *WordNet Search-3.1* (Miller, 1995) was different from other online dictionaries because it showed not only lexical meanings and part of speech, but also its synset relation and word relation. For the purpose of this study, *WordNet Search-3.1* provided word sense observation and selection for both L2 collocates and L2 transferred words as surrogates of L1 collocates.

The use of the two semantic similarity measures, *Adapted Lesk* and *Gloss Vectors*, as a research instrument was operationalized with the online service of *WordNet::Similarity*. In fact, *WordNet::Similarity* conveniently integrated the online service of *WordNet Search-3.1* (Miller, 1995) with hyperlinks and provided semantic similarity calculation by a straightforward process of data input and results output. The process of calculating and retrieving lexical similarity values was as follows.

A. In the context of all word senses:

1. Key-in the L2 collocate in the Word1 slot with the format word#part_of_speech, for example, *observe#v*.
2. Key-in the L2 transferred word from the L1 counterpart in the Word2 slot with the format word#part_of_speech, for example, *celebrate#v*.
3. Select one of the embedded measures, for example, *Gloss Vectors*, with a pull-down menu to calculate the semantic similarity of input words in Word1 and Word2.
4. Press the "Compute" button.
5. Receive the results, e.g, "the relatedness of *observe#v#6* and *celebrate#v#1* using vector is 1". This showed that, among all word sense combination, word sense #6 of *observe* had the highest similarity to word sense #1 of *celebrate*, rated as 1 by the (*Gloss*) *vector* measure.

B. In the context of single word sense: (with the results from the all word sense context)

1. Click on the "View glosses (definitions)" link.
2. Inspect all word senses of the two words and determine a particular word sense for each word.
3. Key-in the L2 collocate in the Word1 slot with the format word#part_of_speech#sense, for example, *observe#v#7* (follow with the eyes or the mind).
4. Key-in the L2 transferred word from the L1 counterpart in the Word2 slot with the format word#part_of_speech#sense, for example, *celebrate#v#1* (behave as expected during of holidays or rites).
5. Select one of the embedded measures, for example, *Gloss Vectors*, with a pull-down menu to calculate the semantic similarity of input words in Word1 and Word2.
6. Press the "Compute" button.
7. Receive the results, e.g, "the relatedness of *observe#v#7* and *celebrate#v#1* using vector is 0.1822". This showed that, for this specific word sense combination, the semantic similarity between *observe* and *celebrate* was rated as 0.1822 by the (*Gloss*) *vector* measure.

The second instrument was a set of collocation candidates were extracted from the collocation lists of previous studies on common miscollocations. The final set of collocation test included two categories of collocation items, congruent verb-noun collocations and incongruent verb-noun collocations. Each category consisted of ten collocation items, as shown in Table 3, with given bases and expected collocates.

3 Verification on the Lexical Similarity Measures

The first quantitative study verified the effectiveness of the two similarity measures, *Adapted Lesk* and *Gloss Vector* based on WordNet, in evaluating semantic similarity. The semantic evaluation test was performed on three sets of ten word pairs. The first set consisted of word pairs that were near synonyms or semantically similar. The second set included word pairs that were semantically related, e.g., they were likely to appear in the same context, but not synonyms. Word pairs in the third set were neither synonyms nor context-related. Table 1 shows the three sets of word pairs designed to manifest differences in semantic distance.

Similar	Related	Unrelated
close, shut	woman, man	door, fish
start, begin	dog, cat	lock, cloth
big, large	tree, leaf	box, eye
end, finish	sun, rain	book, cake
small, tiny	food, eat	bag, road
salary, wage	day, night	brain, cook
injure, harm	body, mind	computer, shoes
grow, raise	animal, human	float, card
exam, test	earth, solar	law, sea
opinion, view	music, melody	color, friend

Table 1. Word Pairs Test Set for Semantic Similarity Measures

The purpose of the semantic evaluation test was to observe how the two similarity measures performed in providing a quantifiable and distinguishable judgment on semantic similarity. Table 2 summarizes statistical descriptions of semantic similarity values calculated by the two similarity measures within the all-word-sense context for each set of word pairs. For each combination of word pair set and similarity measure, the statistical description included mean similarity values and standard deviation in parenthesis.

Type / Measure	Similar	Related	Un-related	Numerical Range
Adapted Lesk	429.6 (358.94)	81.2 (97.41)	19.2 (11.42)	0 → N
Gloss Vectors	0.91 (0.183)	0.46 (0.179)	0.21 (0.072)	0 → 1

Table 2. Statistical Descriptions of Semantic Similarity Values on Semantic Evaluation Test

Note that *Adapted Lesk* was designed to measure semantic similarity with a numerical range from zero to a very large number, with larger number indicating higher similarity. On the other hand, the numerical range of *Gloss Vectors* was normalized to reside in the range of zero and one, with one being the highest similarity. The results showed that both measures of *Adapted Lesk* and *Gloss Vectors* were able to provide reliable and effective indication to semantic similarity and to distinguish semantic relation with the notion of semantic distance. The resulting evidences of the first quantitative study provided support for adopting computational measures of semantic similarity, such as *Adapted Lesk* (abbreviated as *Lesk* in the following sections) and *Gloss Vectors* (abbreviated as *Vector* in the following sections), as objective and systematic method for congruency evaluation.

4 Collocational Congruency Classification by Semantic Similarity Values

As indicated in the previous discussion, current notion of congruency primarily depends on individual researcher's subjective judgment to give a binary classification of congruent and incongruent collocations. This leads to ambiguity as to whether collocations can be consistently classified. Further studies on congruency factor in L2 collocation learning seem to be somewhat problematic in deriving theory with an indeterministic basis. The result of the first study suggested that the values of lexical semantic similarity measures could be considered as effective indicators of the collocation congruency. They are objective and systematic as the numerical values are calculated by computational algorithms and a proper threshold for congruency classification can be ascertained and applied to all evaluation targets, which also lead to consistent classification results.

The second study was designed to demonstrate the application of semantic similarity measures to classification of L2 collocation congruency and to derive empirical results for further deduction. As an operational definition, an L2 collocation is formed by a collocate and a base. In most cases, the base is fully transparent between L2 and L1 and is straightforward to cross-linguistic translation. The collocate, on the other hand, is subjective to cross-linguistic semantic variation and is the sole determinant of congruency. Given an equivalent pair of L2 and L1 collocations, such as "*seek information*" and "尋找資訊 (xyun zhiao zi xyung)", congruency of the L2 collocation is determined by whether the L2 collocate "*seek*" and the L1 collocate "尋找 (xyun zhiao)" are conceptually equivalent or similar at the semantic level. However, current computational similarity measures are designed for two words of the same language. To evaluate semantic similarity between L2 collocate and L1 collocate, a surrogate of L1 collocate in L2 must be used. This L2 surrogate in semantic similarity evaluation with L2 collocate can be represented by one of the synonymous transferred words from L1 counterpart. For example, the transferred word "*find*" of the L1 collocate "尋找(xyun zhiao)" can be used as a surrogate to compute the semantic similarity with the L2 collocate "*seek*".

On the surface, the requirement of an L2 surrogate for an L1 collocate seemed to be a potential shortcoming. There may be several synonymous transferred words eligible as candidates for the L2 surrogate of an L1 collocate. Semantic similarity may vary with the selection of a transferred word as the surrogate, thus, leading to variable congruency evaluation between an L2 collocate and an L1 collocate. A deeper analysis revealed that the use of an L2 surrogate for an L1 collocate was actually advantageous in providing *learner-centered* congruency evaluation. First, the selection of a transferred word for an L1 collocate reflects the L2 lexical knowledge of a learner. Collocational congruency, thus, depends on L2 learners' proficiency level and becomes relevant to learners' individual status. Second, the process of selecting a transferred word involves the activation of conceptual links in learners' cross-linguistic lexical networks, and thus, closely simulates the actual context of learners' L2 collocation use. Third, the decision of a transferred word also incorporates potential L1 influence on individual learners, and thus, embeds the critical factor into congruency evaluation in real context of L2 learning. All these favorable attributes of adopting semantic similarity measures for congruency evaluation provide strong support for better analysis of realistic congruency effects on L2 collocation performance.

The research design of adopting two semantic similarity measures, e.g., *Adapted Lesk* and *Gloss Vectors*, was based on the consideration of providing more evidences of semantic similarity evaluation on a cross-linguistic collocate pair. Evaluation from both measures can be cross-examined for consistency so as to establish larger confidence on the subsequent congruency classification. Experimental results from the first quantitative study partially verified the evaluative consistency of these two measures. The adoption of two semantic similarity measures in the quantitative studies also allowed the construction of a conceptual space of semantic similarity where distribution of semantic similarity values and area of congruency classification can be figuratively observed.

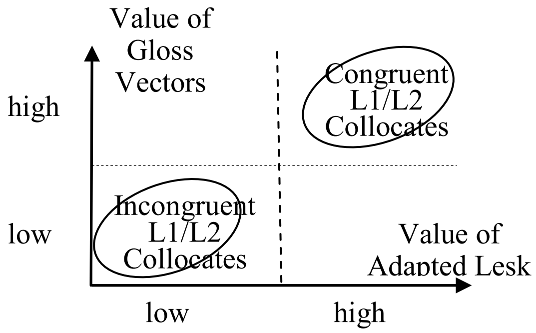

Figure 1. Ideal Semantic Similarity Distribution and Congruency Classification

Figure 1 shows the conceptual space of semantic similarity formed by orthogonal dimensions of the two semantic similarity measures, *Adapted Lesk* and *Gloss Vectors*. It was assumed that incongruent L1/L2 collocates would be given low similarity values from both measures, while congruent collocate pair would receive high values. This would result in an ideal bi-polar distribution of two clearly separated clusters and distinct classification of congruency.

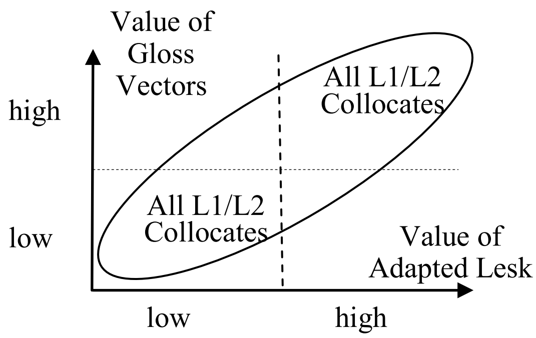

Figure 2. Expectation of Semantic Similarity Distribution

However, this extreme convergence of similarity values may not be realistic. It was expected that some similarity values would fall in the middle range. Thus, actual semantic similarity distribution of cross-linguistic collocate pairs may form a continuous band tilted from lower left corner to upper right end, as shown in Figure 2. In addition, it was conjectured that actual plotting of similarity values of cross-linguistic collocate pairs along with their subjective congruency classification based on human judgment may show an overlapping area.

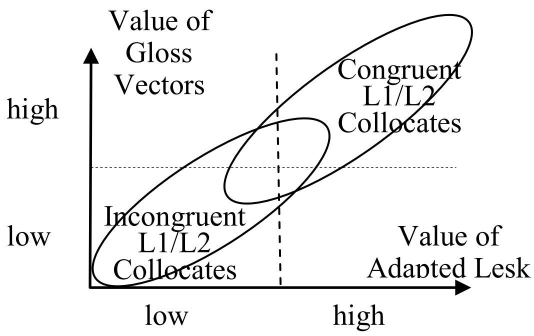

Figure 3. Disagreement between Objective Similarity Evaluation and Subjective Human Classification

This area, as shown in Figure 3, exhibited a boundary crossing disagreement between objective similarity evaluation and subjective human classification where some collocate pairs were humanly judged as congruent but were computationally evaluated as relatively low similarity and some were subjectively incongruent but were objectively of moderate similarity. This mutual middle ground suggested that current practices of subjective human judgment on congruency might actually be partially inconsistent, inaccurate, and unreliable.

5 Applying Sematic Similarity Values and Examining Congruency

The second study applied the semantic similarity measures to the collocation sets so as to provide empirical evidences for the conjectured semantic similarity distribution. Subjective congruency classification of collocations was then cross examined with their computational semantic similarity. Statistical analysis was then performed on the congruency categories for significance of difference in the numerical values of computational semantic similarity. The purpose was to evaluate the consistency of subjective congruency classification from the perspective of objective semantic similarity and to reveal potential classification conflicts.

As noted previously, the congruency classification on the collocation sets was based on an initial and subjective judgment by the researcher. The transferred word from each L1 collocate in the test set was also provided by the participants' most common choice as an exemplar learner's selection. The computation of semantic similarity measures between two words also involves the selection of word sense in two modes. In the single-word-sense mode, each polysemous word was assigned a particular word sense for semantic similarity evaluation. In the all-word-sense mode, no word sense was assigned and all word meanings of the word pairs are considered so as to match the closest word meanings. In other words, the semantic similarity evaluation, when operated in the all-word-sense mode, gives the highest value to represent the most similar word senses of the two polysemous words. In the L2 learning context, semantic similarity evaluation in the single-word-sense mode can be used to simulate lexical knowledge of low-level to mid-level learners, while the all-word-sense mode may assume the characteristics of more advanced learners. When selecting a particular word sense in the single-word-sense mode, L2 learners' primary perception of word meaning would usually be a good consideration.

As introduced previously, the measures of semantic similarity, *Gloss Vectors* and *Adapted Lesk*, provided a deterministic and algorithmic evaluation of semantic similarity between any pair of English words. However, the two measures did not produce a similar range of values. To provide a more convenient and complementary similarity observation of the two measures, the values of the *Adapted Lesk* measure were converted by logarithm, as *Log (Lesk+1)*. The addition of one to the *Lesk* values before logarithmic conversion was to avoid mathematic peculiarity because *Lesk* value started from zero.

For binary classification of similarity level, the thresholds were judiciously ascertained at 0.6 for the *Gloss Vectors* measure and 99 for the *Adapted Lesk* measure (i.e., 2 for *Log (Lesk+1)*). In other words, semantic similarity of a pair of L1/L2 collocates was classified as high if the evaluated value of *Gloss Vectors* measure was higher than 0.6 and/or if the evaluated value of *Log (Lesk+1)* was higher than 2. In a few cases when the evaluative grades were not consistent

Subjective Congruency	L2 Collocate (base)	L1 Transferred Word	Semantic Similarity			
			Single Word Sense		All Word Sense	
			Vector	Lesk	Vector	Lesk
congruent	*acquire* (knowledge)	*get*	1.0	3.364	1.0	3.364
	seek (information)	*search*	1.0	3.018	1.0	3.018
	make (efforts)	*do*	1.0	2.905	1.0	2.905
	see play	*watch*	1.0	2.696	1.0	2.696
	increase (abilities)	*increase*	1.0	3.112	1.0	3.112
	maintain (relationship)	*keep*	1.0	2.560	1.0	2.560
	preserve (culture)	*conserve*	1.0	2.839	1.0	2.839
	make (troubles)	*make*	1.0	2.967	1.0	2.967
	take (actions)	*do*	0.695	2.121	0.695	2.121
	overcome (challenges)	*conquer*	0.750	1.491	0.750	1.491
Incongruent	*surf* (Internet)	*browse*	0.066	1.0	1.0	2.517
	solve (crimes)	*break*	0.731	2.412	0.731	2.412
	make (apology)	*say*	0.421	1.845	0.421	1.845
	study (English)	*read*	0.201	1.462	1.0	2.501
	carry (lanterns)	*hold*	0.141	2.017	1.0	2.605
	ease (worries)	*relieve*	0.249	0.903	1.0	2.220
	make (conclusion)	*get*	1.0	2.967	1.0	2.967
	conduct (heat)	*transmit*	0.083	1.204	1.0	2.452
	make (impression)	*leave*	0.465	1.342	0.583	2.312
	restore (vitality)	*recover*	0.197	1.591	0.197	1.591

Table 3. Semantic Similarity Values of Verb-Noun L2/L1 Collocates in Subjective Congruency Classification

between the two measures, the grade (high similarity or low) given by the *Gloss Vectors* measure was adopted.

Table 3 reports the similarity values of both the subjectively congruent and subjectively incongruent verb-noun L2/L1 collocate pairs in the collocation test set based on the two semantic measures, *Gloss Vectors* and *Adpated Lesk*, in both contexts of single word sense and all word senses.

In Table 3, it was noted that all subjectively congruent L1/L2 verb collocates were indeed of high semantic similarity. In addition, eight out of the ten verb collocate pairs received the highest similarity value in the *Gloss Vectors* measure. The similarity evaluations were also not affected by difference in learners' lexical knowledge as the similarity values were the same in the two modes of single-word-sense and all-word-sense. The experiments showed that subjective and objective evaluations for congruency were consistent on semantically similar pairs of verb collocate. However, most transferred words from participants' common choice were incorrect use. This indicated congruent verb-noun collocations may not be assumed to be easy and straightforward for L2 learners.

For subjectively incongruence, some variations were observed. In the single-word-sense mode, e.g., learners' lexical knowledge on the collocates was assumed to be limited to primary word meaning, most subjectively incongruent verb collocate pairs were indeed of low semantic similarity. The collocate pairs of "*solve* and *break*", "*make* and *get*", were the only two exceptions and surprisingly showed high similarity. In the all-word-sense mode,

Word Sense	Subjective Congruency	N	Gloss Vectors				Log (Lesk+1)			
			Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.	Min.	Max.
Single	Congruent	10	0.945	0.118	0.695	1.0	2.707	0.544	1.491	3.364
	Incongruent	10	0.355	0.305	0.066	1.0	1.674	0.649	0.903	2.967
All	Congruent	10	0.945	0.118	0.695	1.0	2.707	0.544	1.491	3.364
	Incongruent	10	0.793	0.094	0.197	1.0	2.342	0.389	1.591	2.967

Table 4. Descriptive Summary of Computational Semantic Similarity in Verb Noun Collocations

when assumption of learners' lexical knowledge on the collocates was extended to comprehensive word meanings, however, most subjectively incongruent verb collocate pairs showed high semantic similarity. Only three collocate pairs, e.g., "*make* and *say*", "*leave* and *make*", "*restore* and *recover*", remained of low semantic similarity. This empirical results showed that congruency might depend on learners' lexical knowledge on the candidate collocates.

The semantic similarity analysis on verb-noun collocations revealed a problematic pattern of inconsistent classification between human (subjective) and computational (objective) evaluations. This inconsistency of congruency evaluation was further aggravated by the different conditions of learners' word sense level. For verb-noun collocations, the worst inconsistency occurred in the subjectively incongruent category with the assumption of learners' all word senses, where seven out of ten collocations that were humanly judged as of low similarity, were instead, computationally considered as of high similarity.

For further verification, a statistical analysis was also performed on the semantic similarity differences between congruency categories. Table 4 reported the descriptive summary of the computational semantic similarity in verb-noun collocations. Mean values of the *Vector* measures of subjectively congruent collocations, in both contexts of learners' single word sense, and all word sense, was very close to 1.0, indicating that semantics of the L2 collocates and the transferred words from the L1 counterpart were almost identical. indicating that semantics of the L2 collocates and the transferred words from the L1 counterpart were almost identical. For subjectively incongruent verb noun collocations, mean values of the *Vector* measures in the context of learners' single word sense indicated low similarity. However, in the context of learners' all word senses, mean values of the *Vector* measures of subjectively incongruent verb-noun collocations more than doubled and indicated high similarity. A similar pattern of varying similarity evaluation between subjectively congruent and incongruent verb-noun collocations under different contexts of learners' proficiency levels was also observed on the *Lesk* measures.

Table 5 reports the statistical comparison between subjective congruency categories by two computational measures of semantic similarity under learners' different proficiency contexts. It has shown that semantic similarity differences between subjective congruency categories were statistically significant by both measures in the context of learners' single word sense, e.g., F (1, 18) = 32.448, p = 0.000 < 0.05, and F(1, 18) = 14.877, p = 0.001 < 0.05. However, in the context of learners' all word senses, there was no statistically significant difference in the semantic similarity by both measures between congruency categories, e.g., F(1, 18) = 2.228, p = 0.153 > 0.05, and F(1, 18) = 2.984, p = 0.101 > 0.05.

The inconsistency between human (subjective) and computational (objective) congruency classification was manifested in verb noun collocations. Both the item-level and the category-level examination showed that computational and human congruency evaluations might not share the same view. In addition, human congruency evaluation might not account for learners' varying proficiency levels. This analysis revealed that congruency could become ambiguous and disconcerted in the contexts of human evaluation and learners' various proficiency levels. Further studies on better congruency classification and its effects on L2 learners' collocation performance were required.

Word Sense	Subjective Congruency		Sum of Squares	df	Mean Square	F	Sig.
Single	Vector	Between Groups	1.735	1	1.735	32.448	.000
		Within Groups	.962	18	.053		
		Total	2.698	19			
	Lesk	Between Groups	5.334	1	5.334	14.877	.001
		Within Groups	6.454	18	.359		
		Total	11.787	19			
All	Vector	Between Groups	.114	1	.114	2.228	.153
		Within Groups	.924	18	.051		
		Total	1.038	19			
	Lesk	Between Groups	.666	1	.666	2.984	.101
		Within Groups	4.019	18	.223		
		Total	4.685	19			

Table 5. One-Way ANOVA on Computational Semantic Similarity between Subjective Congruency Categories in Verb Noun Collocations

6 Conclusion

The quantitative study empirically verified the applicability of computational semantic measures in classification of L2 collocation congruency. It has shown that objective evaluation of congruency required an input of transferred words from L1 collocate and then operated purely on the L2. This might avoid the fallacy of subjective and cross-linguistic evaluation of congruency. In addition, this *learner-centered* congruency evaluation more closely simulated the context of L2 learners' lexical decision process.

References

B. Wolter and H. Gyllstad. 2011. Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. Applied Linguistics, 32(4), 430-449.

C. Lin. 1997. Semantic network for vocabulary teaching. *Journal of Taiwan Normal University: Humanities & Social Science*, 42, 43-54.

G. A. Miller. 1995. WordNet: A Lexical Database for English. Communications of the ACM (11).

H. J. Chen. 2011. Developing and Evaluating a Web-based Collocation Retrieval Tool for EFL Students and Teachers. Computer Assisted Language Learning, 24(1), 63-79.

J. Yamashita and N. Jiang. 2010. L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. TESOL Quarterly, 44(4), 647-668.

L. J. Brinton and D. M. Brinton. 2010. The lingustic structure of modern language. Philadelphia: John Benjamins.

S. Webb and E. Kagimoto. 2009. The effects of vocabulary learning on collocation and meaning. TESOL Quarterly, 43(1), 55-77.

T. Koya. 2005. The acquisition of basic collocations by Japanese learners of English. Unpublished dissertation. Waseda University.