

Assessing English Learners' Knowledge of Semantic Prosody through a Corpus-Driven Design of Semantic Prosody Test

Moussa Ahmadian (Corresponding author)

Associate Professor, Dept. of English Language and Literature, Faculty of Humanities

Arak University, Arak, Code post: 38156, PO. Box: 879, Iran

Tel: 98-86-1313-5111 E-mail: M-ahmadian@araku.ac.ir; Moussa_ahmadian@yahoo.com.uk

Hooshang Yazdani

Assistant Professor, Dept. of English Language and Literature, Faculty of Humanities

Arak University, Arak, Code post: 38156, PO. Box: 879, Iran

Tel: 98-91-2426-2201 E-mail: H-yazdani@araku.ac.ir

Ali Darabi

Office of Education, Khoramabad, Iran

Tel: 98-91-6663-1982 E-mail: alidarab1350@gmail.com

Received: June 13, 2011

Accepted: July 8, 2011

Published: December 1, 2011

doi:10.5539/elt.v4n4p288

URL: <http://dx.doi.org/10.5539/elt.v4n4p288>

Abstract

This paper introduces a corpus-driven measure as a method to assess EFL learners' knowledge of semantic prosody. Semantic prosody here is defined as the tendency of some words to occur in a certain semantic environment. For example, the verb 'cause' is associated with unpleasant things—death, problem and the like. Subjects were 60 Iranian Persian-speaking English learners drawn from 180 candidates taking English classes in five language institutes. To estimate the quality of the test, a 70-item test of semantic prosody was constructed, validated, and used to measure the subjects' knowledge of semantic prosody. The items were selected from COBUILD Dictionary and were mainly based on those cases of semantic prosody whose conditions (positive or negative) had been already determined by researchers. A proficiency test was applied to determine learners' level of language proficiency as a variable which may affect the results. Data analysis showed that learners' knowledge of semantic prosody is, and can be, appropriately measured by the corpus-driven test of semantic prosody. The implications of the findings for teachers, learners, and test developers are discussed.

Keywords: English language learners, Semantic prosody, Receptive semantic prosody, Productive semantic prosody, Corpus-driven test

1. Introduction

In the last few years, much research has been focused on some specific uses of collocations. Corpus linguists including (Sinclair, 1991), Louw (1993), Stubbs (1995) and Hoey (2003) have provided some instances in which a single word, further to having different collocational behavior, may have different connotations compared with its near synonym (*cause death* but *bring about happiness*). They call this relationship *semantic prosody*. Louw (1993) presents a working definition of semantic prosody as follows:

Semantic prosody refers to a form of meaning which is established through the proximity of consistent series of collocates often characterizable as positive or negative and whose primary function is the expression of the attitude of its speaker or writer toward some pragmatic situation (p.8).

The importance of semantic prosody in language pedagogy has been well recognized by researchers including Sinclair (1991), Louw (1993), Stubbs (1995), and Hoey (2000). Based on their views, teachers, learners, and lexicographers have been advised not to use words with close meanings (near synonyms) at the expense of focusing on connotative meanings (semantic prosodies). It means that as words with denotative meanings usually differ in

their collocational behavior (*substantial meal* but *big food*) and semantic prosodies (*cause death* but *bring about happiness*), the traditional practice of explaining meaning to learners or interpreting meaning to translators and lexicographers should be used with caution.

Based on the above understandings, it seems that most of the existing studies of semantic prosody are confined to the corpus-based description of native speakers' corpora or cross-linguistic studies between native and non-native corpora (McEnery & Xiao, 2006). To date, contrary to well accepted and practiced vocabulary testing, no specific study has been devoted to see whether learners' knowledge of semantic prosody is appropriately recognized, produced, or performed through a corpus-driven test in an EFL context. Therefore, this paper attempts to bridge the gap and supplement the existing studies of semantic prosody. The results may hopefully have implications for EFL teaching and testing.

2. Literature Review

2.1 The concept of semantic prosody

Semantic prosody is a concept developed in linguistic studies. The term has been defined variously by Sinclair (1991), Louw (1993), Stubbs (1995), Hoey (2000), Sardinha, (2000), and Ping-Fang and Jing-Chun (2009). Each definition is basically the same, but the scope of semantic prosody has been expanded by each new definition. Sinclair (1991) noted the fact that certain words seemed to collocate with semantic features of other words that were decidedly either positive or negative. He (1991:112) then states:

Many uses of words and phrases show a tendency to occur in a certain semantic environment. For example, the verb happen (italics original) is associated with unpleasant things, accidents and the like.

However, Sinclair (1991) never came out publicly with the term *semantic prosody* and it was not until 1993 that it was first discussed in details by Louw as a concept in its own right. Louw states that semantic prosody is the "consistent aura of meaning with which a form is imbued by its collocates" (p.157).

Ping-Fang and Jing-Chun (2009: 20) define semantic prosody as "the associative meaning resulting from its collocates and is partially recorded in English Learners' Dictionaries". In Firth's (1957, cited in Ping-Fang and Jing-Chun, 2009: 20) view, the term prosody traditionally refers to "phonological coloring" which goes beyond segmental boundaries. Motivated by Firth, Sinclair (1991) found that "many uses of words and phrases show a tendency to occur in a certain semantic environment" (p. 112), which means that there does exist "some kind of spreading of connotational coloring beyond single word boundaries, which is called semantic prosody" (Partington, 1998: 68).

Furthermore, Ping-Fang and Jing-Chun (2009: 21) argue that "semantic prosody, which is a kind of semantic overflow happening in the syntactic combination, is one specific part of restricted selections, in which a semantic harmony is needed to keep the node words which fulfills the demands of collocates".

Sardinha (2000: 2) also looks at semantic prosody as relating integrally to the connotation of lexical items in a semantic field. In Partington's (1998:68) view, connotational coloring beyond single word boundaries is interpretable in terms of semantic prosody. Zhang and Ooi (2008: 2), similar to Partington's view, define semantic prosody as an "abstract attitudinal, nuanced meaning" or prosody which, in the sequence of the words, colors the selection of the forms. It is inferred from the literature that semantic prosody expresses the *function* of the lexical item (Sinclair, 1991; Stubbs, 2001).

2.2 Corpus-based studies

Based on the above understandings of the concept of semantic prosody, a number of empirical studies have been carried out and presented in the field, the sketch of which is reviewed here. One of these is McEnery and Xiao's (2006) study in which they compared three groups of near synonyms in English with their Chinese equivalents to determine their collocational behavior and semantic prosody, drawing upon data from English and Chinese corpora. Using the statistical test of MI (Mutual Information) to measure collocational strength, they concluded that semantic prosody and semantic preference are as observable in Chinese as they are in English. It was also shown that the semantic prosodies observed in general domains may not apply to technical texts. Furthermore, it was revealed that the collocational behavior and semantic prosodies of near synonyms are quite similar in the two languages. More considerably, this observation echoes the findings which have so far been reported for related language pairs, e.g. English vs. Portuguese (Sardinha, 2000), English vs. Italian (Tognini-Bonelli, 2001), and English vs. German (Dodd, 2000) (all in McEnery & Xiao 2006: 16).

In another cross-linguistic, semantic study, Zhang and Ooi (2008) compared the concept *emotion/feeling* with its Chinese equivalent *qing*; they used two monolingual corpora (Chinese Internet Corpus of 280 million words and

the Bank of English comprising 450 million words) for the analysis of instances of use, and applied Sinclair's lexical model. This model suggests a typical sequence of units of meaning that relates to a lexical item as follows:

Semantic prosody + semantic preference + colligation + collocation + CORE lexical item.

Accordingly, the speaker or writer first selects an abstract attitudinal, nuanced meaning or "prosody" which "colors" the choice of the forms in the sequence; semantic preference refers to the meaning of a group of words that share similar semantic features and "controls" both the colligational and collocational patterns. Colligation has to do with co-occurrence of grammatical choices. It is "one step more than collocation" (Zang and Ooi, 2008: 2). The authors, then, concluded that the Chinese *qing* terms *ganqing/gingan* differ from their English near-equivalents *feeling/emotion* in terms of colligation, collocation, semantic preference and semantic prosody. This model provides a feasible and clear way to accurately grasp the exact meaning of and finer distinctions between the lexical items compared. The study also shows that specific cultural difference affects the nuances of meaning and thus influences semantic prosody.

In the same line, Sinclair (1991) showed that the phrasal verb *SET IN* occurs primarily with subjects that refer to unpleasant states of affairs, such as *rot, decay, malaise, despair, ill-will* and *decadence*. Sinclair (1991: 112) noted that the Lemma HAPPEN "is associated with unpleasant things, accidents, and the like".

However, Stubbs (1995: 25) argues that "although negative prosodies are probably more common, positive prosodies also exist". He provides the example *causing work* which usually means bad news, whereas *providing work* is usually a good thing.

Wang and Wang (2005) examined the semantic prosody of CAUSE. The study showed that great differences exist in the semantic prosody of CAUSE between Chinese learners of English and English native speakers. Chinese learners of English underuse the typical negative semantic prosody and at the same time overuse the atypical positive semantic prosody. However, the study is confined to the semantic prosody of CAUSE without adequate attention to its collocation patterns.

2.3 Testing vocabulary size and depth

Generally, a dichotomy has traditionally been established in the field of vocabulary testing with respect to the nature of lexical competence: the distinction between breadth and depth of vocabulary knowledge (Anderson & Freebody, 1981). The former tries to cover the number of words the students know, i.e. the size of their lexicon (Jaen, 2007), while the latter refers to the degree to which students possess a multidimensional qualitative knowledge of words including pronunciation, spelling, meaning, register, frequency, and grammatical and collocational patterns (Qian & Schedl, 2004).

To investigate categories of lexical depth, measures of collocations have been developed. Collocational measures seem to fall into two categories: the ones which attempt to test productive knowledge and those assessing receptive knowledge. The former was the only aspect investigated during the decade of the nineties, when Bahns and Eldaw (1993), Biskup (1992), and Farghal and Obiedat (1995) designed the first tests of collocations (Jaen, 2007). In the current decade, however, most of the researchers' attention has been focused on the design of the receptive category of the collocation measures (Barfield, 2003; Bonk, 2001; Keshavarz & Salimi, 2007; Mochizuki, 2002). However, testing semantic prosody has not been investigated directly and adequately so far. This paper attempts to consider this point.

3. The study

As mentioned before, the present study tries to assess EFL learners' knowledge of semantic prosody through a corpus-driven test. To do this, the researchers intend to spell out the procedures taken for the study reported below. Based on the aims of the study, the following question was raised: Does the present corpus-driven test of semantic prosody meet the quality of an appropriate test to measure EFL learners' knowledge of semantic prosody?

To answer the above question more objectively, the following null hypothesis was formulated and tested out: The present corpus-driven test of semantic prosody does not meet the quality of an appropriate test to measure EFL learners' knowledge of semantic prosody.

3.1 Participants

The subjects participating in this study were 60 EFL learners (40 male and 20 female) who were randomly selected from among 180 candidates studying English at five English language institutes in Khoramabad, Iran. Their age ranged between 18 and 23 years. They had studied the *Interchange series* for two years and had just entered the *Passage series*, which is a higher level than *Interchange series* and the learners are to know at least 2000 English words. Sex was not considered as a variable in this study. The main reason for choosing these subjects was that they

attended English classes eight terms per year, six weeks per term, and three sessions per week. In other words, they took about 200 hours of English classes for one year. Thus, they had a greater chance to improve their language proficiency.

3.2 Instrumentation

Four different types of instruments were used in this study. The first one was a Michigan Test of English Language Proficiency (1997); it was administered to assess the participants' level of language proficiency. The validity of this test was already presupposed. However, the reliability index, as estimated through Kuder and Richardson formula (KR-21), was reported to be 0.89.

The second instrument was a vocabulary test whose source was *Collins COBUILD Advanced Learner's English Dictionary (2006)* from which the researchers selected the vocabulary items for the development of the semantic prosody test. This new edition updates the snapshot of current English, and contains some attractive features to make the volume even easier to use. One of the features of this Dictionary is that the definitions (or explanations) are written in full sentences, using vocabulary and grammatical structures that occur naturally with the word being explained. Based on the above features, it was felt that it includes the conditions of semantic prosody more than other traditional dictionaries, and thus is appropriate as a source for the purpose of test development.

The third applied instrument was a 70-item *Semantic Prosody Test (hereafter, SPT)* consisting of two sub-tests: *Receptive Semantic Prosody (RSP)* and *productive Semantic Prosody (PSP)*. The information on the reliability and validity of the SPT and its sub-tests will come in the section on pilot study.

The fourth instrument was a validated Criterion Collocation Test (CCT) developed by Chen (2008) to assess the English collocation competence of college students in Taiwan. The CCT is a 50-item multiple choice test containing verb, adjective, and proposition items. The validity of this test was presupposed. This test was run as a criterion measure against which the concurrent validity of the SPT was established. In this study, the reliability estimate of CCT was reported to be 0.81.

4. Procedure

4.1 Item selection and test construction

The items selected for the intended test (SPT) included those cases of semantic prosody whose conditions (positive or negative) had been determined before by different researchers (McEnery & Xiao, 2006; Sardinha, 2000; Sinclair, 1991; Stubbs, 1995; Wang and Wang, 2005; Zhang and Ooi, 2008). Once the items were constructed, they were given to two EFL university lecturers at Arak University (Iran) for their expert comments and advices. They were requested to analyze each item on the basis of its perceptual complexity and face validity.

To that end, a 70-item test was designed, divided into two sub-tests of receptive semantic prosody (40 items) and the productive semantic prosody (30 items). The basic reason for including two sub-tests was to make it possible so as to measure both passive and active knowledge of semantic prosody. The multiple-choice format and the matching items were used for *receptive* tasks. For this task, students were presented with the definitions of the concepts expressed by the target collocations as provided by the *Collins Cobuild English Dictionary (2006)*. An example of an item for multiple-choice receptive tasks is presented below.

(Ex. 1): When people have an **amicable relationship**, they are.....to each other:

- a. enemy b. pleasant c. opponent d. none of these

Finally, as it is seen in the example above, the fourth choice provided in this item, and in each item as well, was "none of these". This alternative, which was the correct answer in 10% of the items, was introduced to minimize the effect of guessing (López-Mezquita, 2005, in Jaen, 2007); this improves test discrimination and reliability (Jaen, 2007).

For the assessment of candidates' *productive* knowledge of semantic prosody, filling-in and translation tasks were used. In this case, this item-response format was closed-ended, and students were asked to complete a definition of the concept expressed by the intended collocations. When these items prompted more than one correct answer, they were all accepted. This was, for example, the case in the following item of the productive SPT, where both "unintelligible" and "abstruse" were accepted:

(Ex. 2): A/An**TALK** is the one you find difficult to understand.

For translation task, however, some incomplete English statements (with their base nouns left out) were presented with their complete Persian translations. The base nouns in Persian were underlined and the subjects were required to fill in the blanks with appropriate English equivalents for the underlined base nouns in Persian. Table 1 shows the SPT content specification.

4.2 Piloting the test

One of the most important stages of the construction of a language test which helps decision-making is piloting that test (Baker, 1989; McNamara, 2000; Bachman, 1990; Bachman and Palmer, 1996). This usually involves administering the test to a known population so that the analysis will throw light on the behavior of the test. Accordingly, in the present study, different steps were taken to collect information about the usefulness of the test itself, and for the improvement of testing procedures. The first step was item analysis. After a set of items for each sub-test was written, reviewed by experts, and revised on the basis of their suggestions, the SPT was ready for trial. To that end, the test was administered to a selected group of 30 EFL learners. A thorough item analysis was conducted in order to obtain the index of item difficulty and item discrimination. The scores collected from this administration were analyzed using Brown's (2004) cut-off score.

The next step in the process of estimating test quality was to calculate the reliability. For this purpose, *Kuder-Richardson formula (KR-21)* was run. This is generally assumed as the best technique to find out inter-item consistency of any test (Brown, 2004; Best & Kahn, 2006). The reliability estimate for SPT was .84 and for receptive and productive sub-tests was estimated to be .82 and .61, respectively.

The last phase of determining test quality was establishing the validity of the test. For this purpose, the researchers applied more than one evidence to support the validity of the SPT: *Internal consistency* and *Concurrent validity*. To satisfy the former, the scores of the sub-tests (receptive and productive) were correlated with each other and also with total SPT. The results (see Table 2) showed high internal consistency between the SPT and its sub-tests. Chen (2008) believes that if the newly developed test is a valid measure of semantic prosody (SP), it will significantly correlate with the outside criterion measure of the same language ability. Based on this idea, and to establish the concurrent validity, the scores on SPT were correlated against those of the criterion collocation test (CCT). The results showed that the test relatively fulfills the criterion of concurrent validity (see Table 3).

4.3 Data collection and data analysis

After fulfilling the requirements of the test construction mentioned above, and before administering the SPT, the Michigan Test (MT) was given to 60 participants to determine their proficiency level as a variable which may affect the results. Thus, to have three proficiency groups, the following steps were taken. Students performing one standard deviation above and below the mean on MT were assigned to the Mid group. Those scoring more than one SD below the mean were assigned to the Low group. Participants who scored more than one SD above the mean entered the High group. In the next step, and within time period of one week, the validated (standardized) SPT was administered to the same target group (60 EFL learners).

During the administration phase of the study, some careful steps were taken. First, attempts were made to seat candidates in an almost stress-free atmosphere for the reduction of test anxiety. However, to enhance the motivation of the subjects so that they could answer the questions honestly and meticulously, they were assured of the confidentiality of the results. In terms of administration and timing for both the SPT and the MT, students were allowed 70 and 100 minutes to complete the tests, respectively. However, most of the subjects were able to finish them before allocated time. This would indicate that the measures were correctly designed or chosen from a practical point of view. Moreover, items were designed (even for fill-in and translation tasks) in objective formats. Therefore, there was no problem of inter-rater reliability. Correct answers scored one point and incorrect answers scored zero.

As for data analysis procedures, some statistical measurements were applied: For establishing the reliability of the SPT and its sub-tests, Kuder and Richardson (KR-20) formula was used. To fulfill the requirement of validity (internal consistency and concurrent), *Pearson Correlation* analysis was applied. Furthermore, the statistical measurement of One-Way ANOVA was used to compare the participants in terms of their performance on both the Michigan test and the Semantic Prosody Test.

5. Results

As mentioned before, the SPT was analyzed for its appropriateness in terms of item characteristics (item difficulty and item discrimination) and test characteristics (reliability and validity). As shown in Table 4, after an analysis of item difficulty, 3 items (all of them belonging to the PSP) obtained p-values of .0, since they prompted incorrect answers from all the participants. As expected, the discrimination index showed that these highly difficult items were non-discriminating among candidates, and so they would need to be replaced in future studies by more relevant items. The rest of the numerical values yielded by the item difficulty analysis were classified following Ebel's (1965, cited in Cervantes, 1989) criteria (Table 4.): 12 items (16%) were classified as very difficult, 21 items (30%) as difficult, 28 items (40%) offered a desirable level of difficulty, 8 items (9%) were easy and finally 1 item (1%) fell into the

category of very easy items.

In order to obtain the reliability coefficient, we ran Kuder and Richardson formula for total scores and for the sub-tests individually. The internal reliability values found in the RSP (.82) as well as in the PSP (.61) were relatively acceptable. However, taking the test as a whole, its overall reliability was highly acceptable (.84). As for validity, one criterion used was internal consistency. Based on the correlational indices presented in Table 2, the correlation coefficient between two sub-tests (RSP and PSP) is .45 which is considered as moderate. Between PSP and total SPT, the value is .74 which is stronger than the first one. However, the correlation coefficient between RSP and total SPT is .92 which is the strongest, compared to the other two correlations. These correlational matrices show that the relationship between the variables are all significant, indicating that the SPT and its sub-tests are internally consistent.

Another criterion applied to determine the validity of SPT was concurrent validity. To fulfill this requirement, the scores on SPT and CCT (criterion measure) were correlated (see Table 3). Obtained significant correlation coefficient (.289), though low, was an evidence to support the intended concurrent validity.

In this study, the result of learners' performance on SPT was worth elaborating. However, before this analysis, it was needed to elaborate on the results of MT scores. The descriptive statistics for proficiency scores (see Table 5) shows that the mean of scores on this measure is 60.02 and SD is 14.269. Moreover, the scores on this test ranged from 33 to 86 showing a great division between the minimum score and the maximum one. The information concerning SD (14.269) shows a high normal distribution in MT scores.

As mentioned in the procedure section, the mean and standard deviation on the MT scores were used as the criteria for determining the proficiency levels of the participants. However, to be sure of group difference with regards to proficiency level, a One-Way ANOVA (Table 6) was run to examine the differences in group means. An F value of 86.053 at 0.000 level of significance is observed which verifies that the test is able to differentiate groups with different proficiency levels.

Concerning the results of learners' performance on SPT, the analysis (see Table 7) showed that the mean of correct answers in the whole test (i.e. including both sub-tests) was 29.63%, a considerably low score. Furthermore, the standard deviation (S.D.) is 9.36, which is relatively low showing that the group is fairly homogeneous in their level of collocation knowledge of semantic prosody. Moreover, from a comparison between data obtained from both sub-tests (Table 7), we observe a clear difference between the mean scores in the RSP (21.75) and the PSP (7.72) subtests. Equally interesting was the information concerning SD in both sub-tests. Oddly enough, subjects' scores were more uniform in the PSP (3.76) than in the RSP (7.11) sub-tests.

Finally, a One-Way ANOVA was run to determine whether level of language proficiency makes any difference among EFL learners in terms of their performance on semantic prosody test. The result showed an F- value of 3.084 at .05 level (Table 8) which was not significant. In other words, the means of proficiency groups (High, Mid, and Low) on the semantic prosody test were not significantly different from each other.

6. Discussion

The results of statistical analyses for test validation showed that reliability coefficient of the whole SPT was satisfactory, going beyond .8—a conventional yardstick against which reliability is measured (Jaen, 2007:140). This satisfactory result may be attributable to the careful and systematic corpus-driven design, and perhaps, to the construction of the test items. It also holds true for the RSP sub-test in which the reliability coefficient goes beyond the specified yardstick. However, for PSP sub-test, the reported reliability coefficient is less satisfactory. It is considered that this less satisfactory estimate of reliability for PSP was due to the small number of items (30) and the little variance existing among subjects' performance.

As for validity, the coefficient of correlation between SPT and its sub-tests were significant ($P < 0.05$). An inspection of the results of coefficient of internal consistency shows that SPT demonstrates lower correlation to PSP than to RSP sub-test. This may be due to difficult nature of productive items evidenced in Jaen's (2007) study in which he concluded that learners have more problems with producing collocations than with recognizing them. Concurrent validity as another evidence for estimating the quality of the present test was reported to be low, though significant. This may be possibly due to the discrepancy of purposes between the SPT and the criterion collocation test. It can also be said that the SPT and the CCT do not measure the same general area of behavior or they may not have the same name. These explanations are supported by what Bachman (1990) purports. According to him, some correlations, if moderately high, can be cited as evidence that the new test measures approximately the same general area of behavior as other tests designed by the same name as the new test. The correlation results of the present study are also in accordance with the theoretical assumption of Murphy and David Shofer (1998 in Miao, 2006), that

is, theoretically a correlation could range in absolute value from 0.0 to 1.0, whereas in practice, most validity coefficients tend to be fairly small. A good, carefully chosen test is not likely to show a correlation greater than .50 with an important criterion and, in fact, validity coefficients greater than .30 are not all common in applied settings (Miao, 2006).

Not contrary to the above justifications, Hatch & Farhady (1982) pinpoint that in interpreting a variable we should depend more on logical reasoning rather than on figures. "A correlation coefficient may be very high but meaningless, or it may be fairly low and still meaningful" (Hatch & Farhady, 1982:208). It is important to note here that any interpretation depends on what variables are being compared and what kind of decisions must be made on the basis of the discovered relation.

Whatever the results of the estimation of test quality were in this study, the descriptive statistics on SPT and its sub-tests showed that the overall performance of the learners on the SPT was weak. This may be possibly due to the fact that a big challenge in learning a word lies in mastering its pragmatic function (Zhang, 2008), which is related to its semantic prosody (Partington, 1998; Zhang, 2009). This finding does not run counter to Nesselhauf's (2003) contentions that collocations have been largely neglected by researchers, course designers, and EFL practitioners. Accordingly, researchers like Zughoul & Hussein (2001) and Keshavarz & Salimi (2007) found that EFL learners have insufficient knowledge of English collocations, thus their findings are proved to be in line with the present study. More importantly, it was shown that knowledge of semantic prosody seems to be more difficult at the productive than at the receptive levels, a finding which empirically confirms the generally held hypothesis that this type of combination is particularly problematic for students in their linguistic production (see Jaen, 2007). The information concerning SD in both sub-tests of RSP and PSP showed that subjects' scores were more uniform in the PSP than in the RSP. One possible explanation for this could be that RSP discriminated between high and low level candidates while the PSP produced such low scores that no variance was observable or it is better to say all candidates showed the same lack of knowledge. Some similar results were reported in Jaen's (2007) study concerning the analysis of his subjects' performance on receptive and productive sub-tests of collocation behavior.

Seeking other possible explanations for the subjects' poor performance on SPT, the researchers feel, and it might be the case, that most of the monolingual dictionaries from which learners get benefit have no or poor information on the conditions of semantic prosody, thus allowing learners not to be familiar with such uses and conditions. One more point to consider is that, based on the results, the means of proficiency groups (High, Mid, and Low) on the semantic prosody test were not significantly different from each other. Thus, it is likely that knowledge of semantic prosody is neglected by the least and the most proficient L2 learners almost equally well, indicating that level of language proficiency does not have any possible effect on semantic prosody.

By and large, it should be said that though the way we measured learners' knowledge of semantic prosody through recently developed corpus-driven test was novel in its direction and unique in its scope, the measurement device developed may not be so satisfactory in terms of the criteria of test quality (at least for concurrent validity). It is advisable to improve this by more in-depth processes of test construction. For future studies, this may be done by selecting and including more cases or samples of semantic prosody, further to other issues relevant and essential to test item construction. It should be further noted that though it is in its early stages of development, the prospect of corpus-driven SP test construction seems to be encouraging and fruitful.

7. Conclusions, Implications and Suggestions

The analysis of the SPT carried out among EFL learners led to some conclusions. First, based on the results, it can be concluded that though knowledge of semantic prosody is considered to be undermined (by most EFL learners) in receptive and, to a great extent, in productive modes, the present corpus-driven test of semantic prosody is of modest reliability and validity. From this finding, it can further be concluded that careful and systematic selection of the items, not specifically based on intuition and word frequency, might contribute to test quality as well as test usefulness.

It can also be concluded that learning individual words and their meanings does not suffice to achieve great fluency in a second language. Knowing the way words combine into chunks (collocations) characteristic of the language, as well as being aware of the conditions of semantic prosody is necessary. Moreover, it should be noted that from the very beginning, learners' attention should be turned to these kinds of combinations (words) and conditions (semantic prosody), and students should be constantly acquainted with an increasing number of collocations, and eventually the learners' progress in SP should be measured accordingly.

The findings of this study can have some implications too. First, taking benefits from the findings of the present study, teachers can realize the significance of semantic prosody in ESL/EFL learning and teaching (Partington, 1998; Hoey, 2000; Nesselhauf, 2003; McEnery & Xiao, 2006). Second, by constructing tests of this kind, teachers can motivate

the learners to move in this productive use of language. Awareness of semantic prosody can be greatly beneficial in helping language learners understand how to use lexical items appropriately. In this study, learners showed insufficient knowledge of semantic prosody; this insufficiency is reflected in the test recently constructed. Test developers can also benefit from the systematic selection, construction, and development of test items shown in the present study and follow the same procedures when devising tests for their professional purposes.

This study may also motivate interested researchers to conduct further research on this issue. For example, conducting a corpus-based research is needed in order to explain the degree to which the conditions of semantic prosody (positive, negative, and neutral) have been used in the learner corpora and compare these conditions with those of native corpora and construct relevant tests. Still further studies with the cross-linguistic analysis of the use of unusual semantic prosody (irony, for example) in both English and Persian may produce more interesting results. Finally, more corpus-driven research is needed in order to analyze the degree to which the interpretation on semantic prosody is influenced by syntactic representation.

References

- Anderson, R., & Freebody, P. (1981). Vocabulary knowledge. In J.T. Guthrie (Ed.), *Comprehension and teaching: Research reviews*. Newark, DE: International Reading Association. pp. 77-117
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L., & Palmer, A. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, 21, 1, 101-114. [http://dx.doi.org/10.1016/0346-251X\(93\)90010-E](http://dx.doi.org/10.1016/0346-251X(93)90010-E)
- Baker, D. (1989). *Language testing: A critical survey and practical guide*. London: Edward Arnold.
- Barfield, A. (2003). *Collocation recognition and production: Research insights*. Tokyo: Chuo University.
- Best, J. W., & Kahn, J. V. (2006). *Research in education* (10th ed.). New York: Pearson Education Inc.
- Biskup, D. (1992). L1 influence on learners' renderings of English collocations. A Polish/German empirical study. In P. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics*. London: Macmillan. pp. 85-93
- Bonk, W. J. (2001). Testing ESL learners' knowledge of collocations. In T. Hudson & J.D. Brown (Eds.), *A focus on language test development: expanding the language proficiency construct across a variety of tests. (Technical Report #21)*. Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center. pp. 113-142
- Brown, H. D. (2004). *Language assessment: principles and classroom practices*. New York: Longman.
- Cervantes, E. (1989). Designing a reading and listening test for a specific purpose. *English Teaching Forum*, 27 (1), 10-14
- Chen, M. H. (2008). A study of English collocation competence of college students in Taiwan. *Unpublished Master's Thesis*. Department of Applied Foreign Languages, National University of Taiwan, Taipei, Taiwan.
- Farghal, M., & Obiedat, H. (1995). Collocations: A neglected variable in EFL. *International Review of Applied Linguistics*, 33(4), 315-331. <http://dx.doi.org/10.1515/iral.1995.33.4.315>
- Hatch, E., & Farhady, H. (1982). *Research design and statistics for Applied Linguistics*. Cambridge, MA: Newbery House.
- Hoey, M. (2000). A world beyond collocation: New perspectives on vocabulary teaching. In M. Lewis (ed.), *Teaching collocations: Further developments in the lexical approach*. Boston: Heinle. pp. 224-245
- Hoey, M. (2003). Lexical priming and the qualities of text. [Online] Available: <http://www.monabaker.com/tsresources/Lexical Priming and the Properties of Text.htm>. (October 14, 2008)
- Jaen, M. M. (2007). A corpus-driven design of a test for assessing the ESL collocational competence of university students. *International journal of English Studies*, 7 (2), 127-147
- Keshavarz, M. H., & Salimi, H. (2007). Collocational competence and cloze test performance: *A study of Iranian EFL learners*, 17(1), 81-92
- Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, & E. Tognini-Bonelli (eds.), *Text and technology: In honour of John Sinclair*. Amsterdam: John Benjamins. pp. 157-176
- McEnery, A., & Xiao, Z. (2006). Collocation, semantic prosody and near synonymy: A Cross-linguistic perspective. *Applied linguistics*, 27(1), 103-199. <http://dx.doi.org/10.1093/applin/ami045>

- McNamara, T. (2000). Language assessment as social practice: Challenges for research. *Language Testing*, 18 (4), 333-349
- Miao, Y. (2006). Validating a simulated test of CET. *Asian EFL Journal Professional Teaching Articles*, 12(4), 1-18
- Mochizuki, M. (2002). Exploration of two aspects of vocabulary knowledge: Paradigmatic and collocational. *Annual Review of English Language Education in Japan*, 13, 121- 129
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24 (2), 223-242. <http://dx.doi.org/10.1093/applin/24.2.223>
- Partington, A. (1998). *Patterns and meanings: Using corpora for English language research and teaching*. Philadelphia, PA: John Benjamins.
- Ping-Fang, Y., & Jing-Chun, C. (2009). Semantic prosody: A new perspective on lexicography. *China Foreign Language*, 7(1), 20-25
- Qian, D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing*, 21, 1, 28-52. <http://dx.doi.org/10.1191/0265532204lt273oa>
- Sardinha, T. B. (2000). Semantic prosodies in English and Portuguese: A contrastive study. *Cuadernos de Filología Inglesa*, 9 (1), 93-110
- Sinclair, J. (1991). *Corpus, concordance, and collocation*. Oxford: Oxford University Press.
- Stubbs, M. (1995). Collocations and semantic profiles: On the cause of trouble with quantitative studies. *Functions of Language*, 2 (1), 23-55
- Stubbs, M. (2001). *Words and phrases*. Oxford: Blackwesi.
- Wang, H., & Wang, T. (2005). A contrastive study on the semantic prosody of CAUSE. *Modern Foreign Language*, 28(3), 297-307
- Widdowson, H. G. (2007). *Discourse analysis*. Oxford: Oxford University Press.
- Zhang, R., & Ooi, B. Y. (2008). A corpus-based analysis of 'qing': A contrastive–semantic perspective. *Proceedings of the International Symposium on Using Corpora in Contrastive and Translation Studies*, Hangzhou, China, September, 25-27
- Zhang, W. (2008). In search of English as foreign language (EFL) teachers' knowledge of vocabulary instruction. *Unpublished doctoral dissertation*. Georgia State University.
- Zhang, W. (2009). Semantic prosody and ESL/EFL vocabulary pedagogy. *TESL Canada Journal*, 26 (2), 1-12
- Zughoul, R., & Hussein, A. (2001). *Collocational competence of Arabic speaking learners of English: A study in lexical semantics*. ERIC Document Reproduction.

Table 1. Content specification of SPT

Components	No. of items	Points
Part I: receptive		
Section A: classifying	10	15
Section B: multiple-choice	20	20
Section C: matching	10	15
Part II: productive		
Section A: gap-filling (without contextualization)	10	10
Section B: gap-filling (contextualized)	10	15
Section C: translation	10	15
Total	70	90

Table 2. Internal consistency of SPT and the sub-tests

		SPT	RSP	PSP
SPT	Pearson correlation	1	.92 (**)	.74 (**)
	sig.(2-tailed)	.	.000	.000
	N	60	60	60

** Correlation is significant at the 0.05 level (2-tailed).

Table 3. Correlation between SPT and CCT

		SPT	CCT
SPT	Pearson correlation	1	.289(**)
	Sig. (2- tailed)	.	.093.
	N	60	60
CCT	Pearson correlation	.289(**)	1
	Sig. (2- tailed)	.093.	.
	N	60	60

** Correlation is significant at the 0.05 level (2-tailed).

Table 4. Analysis of difficulty of individual items according to Ebel's criteria.

	Overall	RSP	PSP
Non-discriminating items (p value = .0)	3 (4%)	0 (0%)	3 (10%)
Very difficult items (p-values from .01 to .14)	12 (16%)	0 (0%)	12 (41%)
Difficult items (p-values from .15to .39)	21 (30%)	8 (20%)	10 (33%)
Desirable items (p-values from .40 to .70)	28 (40%)	24 (62%)	4 (13%)
Easy items (p-values from .71 to .85)	8 (9%)	7 (17%)	1(3%)
Very easy items (p-values from .86 to 1)	1 (1%)	1(1%)	0 (0%)

Table 5. Descriptive statistics for MT scores

N	valid	60
	Missing	0
Mean		60.02
Std. Deviation		14.269
Variance		203.610
Minimum		33
Maximum		86

Table 6. One Way ANOVA for Mean Difference of the MT

	Sum of Squares	df	Mean Squares	F	Sig.
Between Groups	7935.232	2	3967.616	86.053	.000
Within Groups	1198.768	57	46.106		
Total	9134.000	59			

Table 7. Descriptive statistics for SPT and its Sub-tests (Multiple modes exist)

	N	Mean	S.D	Variance	Range	Min	Max
SPT	60	29.63 (42.32%)	9.36	87.626	40	9	49
RSP	60	21.75 (54.37%)	7.113	50.597	30	6	36
PSP	60	7.72 (%25.73%)	3.769	14.206	17	2	19

Table 8. One Way ANOVA for Mean Difference of the SPT scores

	Sum of Squares	df	Mean Squares	F	Sig.
Between Groups	534.973	2	267.487	3.084	.063
Within Groups	2254.889	57	86.726		
Total	2789.862	59			