

2017

Measuring Productive Semantic Associational Knowledge of the Most Frequent Words

Alireza Barouni Ebrahimi
Western University, abarouni@uwo.ca

Follow this and additional works at: <https://ir.lib.uwo.ca/edupub>



Part of the [Education Commons](#)

Citation of this paper:

Barouni Ebrahimi, Alireza, "Measuring Productive Semantic Associational Knowledge of the Most Frequent Words" (2017). *Education Publications*. 276.
<https://ir.lib.uwo.ca/edupub/276>

Measuring Productive Semantic Associational Knowledge of the Most Frequent Words

Alireza Barouni Ebrahimi
University of Western Ontario, Canada
abarouni@uwo.ca

Abstract

Associational vocabulary knowledge is associated with writing and speaking skills. These skills are essential for EAP students who express themselves in oral presentations or written assignments. Therefore, diagnostic measurement of associational vocabulary knowledge is of vital importance, especially in regard to the most frequent 1,000 word families that cover 81% and 85% of written and spoken text. This study measured 46 Iranian EAP students' productive semantic associational knowledge of words at the 1,000 word frequency level. The findings indicate that while participants had productive form-meaning knowledge of the words, they did not seem to have extensive semantic associational knowledge of the same words. This assists in diagnosing area of weakness and the degree to which instructional emphasis on high frequency words might improve their knowledge.

Keywords: Semantic Associational Knowledge; Productive Semantic Associational Knowledge; Productive Semantic Associational Assessment; Measuring Productive Semantic Associational Knowledge; Most Frequent Words

Introduction

Lexical knowledge is related to success in reading, writing, general language proficiency and academic achievement (Laufer, Elder, Hill & Congdon, 2004; Milton, 2013; Milton, Wade, & Hopkins, 2010; Schoonen, 2010). Meara (1996) and Schmitt (2010) argue that vocabulary knowledge makes a significant contribution to almost all aspects of second language proficiency. Among all words, the ones used frequently in a wide range of spoken and written texts are essential in effective comprehension and communication. The most frequent 1,000 word families reoccur so often in spoken and written texts that they have a much greater value for comprehension and use than the second and the third most frequent ones (Webb & Chang, 2012).

Nation's (2001, 2013) conception of vocabulary knowledge includes not just meaning but various other components including associations. Associations help language learners to understand the full meaning of words and recall the word forms or meanings in contexts (Nation, 2013). Therefore, it is important to understand to what extent university students who have studied English for Academic Purposes (EAP) have productive associational vocabulary knowledge of the most frequent 1,000 word families. The current study measures knowledge of semantic associations and corresponding form-meaning connection of the most frequent words.

Literature Review

Associations

The properties of the associations can be categorized in a number of ways, but the most common categories, especially in earlier studies, are paradigmatic, syntagmatic, and clang

(Fitzpatrick, 2006). Meara (1983) asserts that syntagmatic association forms an obvious sequential link with the stimulus word (*bark, spotted, and bite for dog*), paradigmatic association forms the same grammatical form class as the stimulus word (*cat, wolf or animal for dog*), and clang associates are responses which are related to certain phonological features of the stimulus word which carry no clear semantic relationship to it (*save for cave or auto, tomorrow, and zero for borrow*).

In defining syntagmatic, paradigmatic, and clang categories, Fitzpatrick and Munby (2014) refer back to Nation's (2013) aspects of word knowledge – form-based, meaning-based, and use-based knowledge. They point out that clang responses are form-based in that they have phonological similarities to the stimulus words, paradigmatic responses are meaning-based as they are from the same word class and have related meanings, and syntagmatic responses are use-based because they are usually found beside the stimulus word in a text.

Nation (2013) asserts that the associations of a word are mainly the result of the various meaning systems that the word fits into — synonyms, antonyms, family members of the same general headword, words in a part-whole relationship, superordinate and subordinate words. Nation's (2013) framework for productive mode of associations aspect indicates “what other words could be used instead of this one?” (p.49), and Nation (2013) interprets this question as “being able to produce synonyms and opposites for [the target word]” (p. 50). To test this aspect, Nation (2013) asks whether the learner can “recall this word when presented with related ideas” (p.538), and for the purpose of testing specifically this aspect, he asks whether the learner can “add to this list of associated words” (p. 551). Considering the fact that his framework informs the current study, a synonym and antonym test for a part of the purpose of the current study is necessary in the first place. In addition, to

fully serve the purpose of measuring associations productively, literature needs to be investigated to see how associations (related words) were measured productively by different researchers.

Schmitt's and Meara's (1997) instrument had 20 target words followed by three blanks to write three possible associations. A four-point Likert scale (0–3) was also attached to each target word to enable the participants to indicate how well they knew the words. They found that their participants could supply approximately 50% of possible associations for target words they rated as known.

Schmitt (1998, 1999) asked his participants to give 3 responses to each target word. He elicited associations by using the following instructions, "Please give the first 3 words you think of when you hear the word _____." Schmitt (1998) found that the four participants' mean scores of associations were 1.09, 1.64, 1.09, and 1.18 (less than 50%) out of a possible 3. Schmitt (1999) found that of 136 correct associative responses to the TOEFL items, only 45 (33%) matched with native-like associations.

Wolter (2001) measured word associations productively in a part of his study to compare lexical knowledge of 9 English L1 speakers and 13 English L2 learners. For the purpose of his study, he used an aural-written test with the following instructions:

The following test is a word test. You will hear several words and you will be asked to respond with the first word that comes to mind upon hearing the word. There are no right or wrong answers, so try not to take a long time considering your response.

(Wolter, 2001, p.51)

The results showed that while English L2 participants did not provide as many paradigmatic associations as English L1 participants, they provided almost the same number of syntagmatic associations. The provided associations were rated and assigned to the defined categories with two judges. The categories were defined in advance.

In a series of different studies, Webb (2005, 2007a, 2007b, 2007c, 2009) measured associations, as a part of his studies, to investigate the effects of: (a) receptive and productive learning tasks on vocabulary knowledge (2005), (b) contextualized and decontextualized learning tasks on vocabulary knowledge (2007a), and (c) receptive and productive learning of word pairs on vocabulary knowledge (2009b). He also investigated the effects of: repetition (1, 3, 7, and 10 encounters) on vocabulary learning (2007b), and synonymy on vocabulary learning (2007c).

Webb, in his series of studies, had a test of productive knowledge of syntagmatic association in which the participants were asked to write an L2 syntagmatic associate beside each target word. He also had a test of productive knowledge of paradigmatic association in which the participants were required to write a paradigmatic associate beside each target word. Coordinates, superordinates, subordinates, antonyms, and synonyms were all scored as correct in his test. Syntagmatic associates were not scored as correct because the previous test measured this aspect, and this issue was carefully explained in the instructions.

Following Webb's series of studies (2005, 2007a, 2007b, 2007c, 2009), one could conclude that in addition to synonyms and antonyms, coordination, superordination and subordination could also be considered when associations are measured. However, Webb's findings do not seem to inform the current study as he measured the participants' performance on disguised words after his treatments.

Form and meaning

It is believed that form and meaning link is the basic dimension of vocabulary knowledge and most probably the first to be acquired (Laufer & Goldstein, 2004) because words are units of meaning and they play the most important role in communication. Nation (2013)

states that learners consider a word known as soon as they know its form (what the word sounds like or looks like) and its meaning. However, he believes that not only do language learners need to know the form and meaning of a word, but also they need to know how to connect these two.

To measure form and meaning connection, Nation (2013) asks whether the learner can “produce the appropriate word form to express this meaning” (p.538), and for the purpose of testing this aspect specifically, he asks whether the learner can “translate these words into L2” (p. 551).

Laufer and Nation (1999) validated a vocabulary size test of controlled productive ability that is, in fact, a way of measuring form and meaning. For each item in their test, they presented a meaningful sentence context and the first letters of the target item. The first letters prevent the participants filling in another word which could be used semantically in the provided context. Their participants’ mean scores for the 2,000 word level was 11.8 (65.55%), 15 (83.33%), 16.2 (90%), and 17 (94.44%) out of 18 (18 items).

Following Laufer’s personal advice and taking into account Schmitt et al.’s findings, Llach and Espinosa (2014) combined the PLT Version A and Version C (equivalent versions) from the parallel Version 1 (Laufer & Nation, 1999) to have a 30 item test. Their 197 EFL Spanish participants’ mean scores of the 2,000 most frequent words were 7.26, 8.78, and 12.30 out of 30 in three successive grades (1st, 2nd, and 3rd grades of secondary education).

Zheng (2009) administered the VLT and PLT to 88 Chinese university students. The mean score of her participants for the 2,000 word level was 15.61/18 (86.74%), indicating that her participants seemed to have mastery over the words at 2,000 level.

High-frequency words

Words can be broken into 1,000 frequency levels (word families) shown as K1, K2, K3, etc. Nation (2013) categorizes words as high-frequency, mid-frequency, low-frequency and specialized words (including academic and technical words). The high-frequency words include function and content words which contain around 2,000 word families. Nation (2013) notes that this number has become a matter of debate as Schmitt and Schmitt (2012) argue for having a 3,000 word family high-frequency vocabulary list.

The most frequent 1,000 word families reoccur so often in spoken and written text (including academic text) that they have a much greater value for comprehension and use than the second and third most frequent 1,000 word families (Webb & Chang, 2012). The most frequent 1,000 word families cover 81% of written text and 85% of spoken text (Nation, 2006).

Brezina and Gablasova (2013) assert that although there are a number of different lists for English frequent lexical items, the most influential and widely used one is the General Service List (GSL, West, 1953), which has been adopted in pedagogical practice and vocabulary research, and also has served as the non-academic baseline of the Academic Word List (AWL, Coxhead, 2000, 2011). In other words, West's GSL influences directly the way essential English vocabulary is conceptualized and also lies at the center of the distinction between general and academic vocabulary.

However, West's GSL has been criticized over the years mainly for being out-of-date (Carter, 2012; Nation, 1990; Richards, 1976). GSL was also criticized for the combination of objective and subjective criteria on which the wordlist was based (Gilner & Morales, 2008). In other words, Brezina and Gablasova (2013) argue that GSL's compilation involved a number of principles that brought subjectivity into the final product.

In response to the problems identified with the GSL, Brezina and Gablasova (2013) offered an objective approach to the development of the New General Service List (new-GSL) by means of examining frequent general words across a variety of language corpora. In addition, while the GSL is organized according to the word family principle (a headword, its inflected and closely related derivative forms, Nation, 2013), the new-GSL's lexical units are lemmas (a headword, its inflected and reduced forms, Nation, 2013).

The present study

To address the focus of the study, and considering the fact that the available participant pool for the current study was a group of Iranian EAP students, the following research questions guided the design of the study:

1. To what extent do (a group of Iranian) EAP students have productive knowledge of semantic associations at the 1,000 word frequency level?
2. To what extent do (a group of Iranian) EAP students have productive knowledge of the corresponding form-meaning connection at the 1,000 word frequency level?

Method

Design (Associations)

To measure associations, the aforementioned research studies (discussed in Literature Review) have one factor in common. They all present target words and then ask for the words that come to mind as a response (either one word or three words). However, this method of measuring associations seemed to have two problems. First, the participants might provide some collocations of the target words, and this was not the intended response type.

Instead, it was planned to measure the same participants' knowledge of collocations as a separate aspect of vocabulary knowledge with a separate test. Second, the participants would not necessarily provide synonyms and antonyms of the target words. This was a limitation of the test because, based on Nation's (2013) framework, it was intended to measure synonymy and antonymy as a part of measuring the associations (a test was required to be adopted for this specific purpose though).

To address the aforementioned problems, measuring associations in a controlled way—control of producing associations for the test—seemed promising. It was noticed that coordinates, superordinates, subordinates, collocations, synonyms, and antonyms could be potential associations for measurement (see Webb, 2005). Collocations were planned to be measured separately as one of the aspects of vocabulary knowledge mentioned by Nation (2013). Therefore, to have a controlled way of measuring other associations, the intention was to design one test to measure synonymy and antonymy, and another test to measure other forms of associations. As a result, two tests were adopted. Firstly, the following synonym and antonym test was considered:

For the following words, write down a synonym and an antonym.

For example:

	Synonym	Antonym
FRIEND	buddy	enemy
REAL	actual	unreal
HOLD	keep	release

For the purpose of testing other associations in a controlled way, three relationships were considered—coordination, superordination, and subordination. Coordination was, in fact, an umbrella term for all synonymy, antonymy, collocation, superordination, and

subordination. Because synonymy, antonymy, and collocation were planned for separate measures, the focus needed to be on superordination and subordination. In an attempt to measure these two aspects productively, a test with the following instructions was adopted:

For the following words, write down a possible main category and two other examples of the same category.

For example:

	POSSIBLE CATEGORY	EXAMPLE	EXAMPLE
RED	color	blue	white
PIZZA	food	burger	spaghetti
SHIRT	clothes	hat	skirt

There seemed to be a problem though. Words such as *adjective*, *noun*, or *verb* as the possible category of target words could be provided by participants. As a result, the instructions of the test were revised as follows: “For the following words, write down a possible main category and two other examples of the same category. Do **NOT** write *noun*, *adjective*, *verb* or *adverb* for the possible categories”.

Design (Form and Meaning)

To measure form and meaning aspect productively, Nation’s (2013) framework requires the translation of words into L2. Considering the potential participant pool for the current study, Persian to English word translation tests were to be provided. To design the form and meaning tests, the Persian translations of the core meaning (usually the first entry in dictionaries) of each target word were provided. For this purpose, two English to Persian dictionaries were consulted. In other words, I compared what I knew as the translation of the core meaning of words with the translations provided by the aforementioned dictionaries. If there was an inconsistency between what I believed was the core meaning and what

dictionaries showed as the core meanings, the COCA was consulted to judge which meaning was the core meaning. For this purpose, the first 100 examples of the corpus were consulted to select the most frequent meaning of the target word. To help guide the task and to prevent the participants from providing possible synonyms for the target words, letters of each target word presented by dashes and one letter in the target word (neither the beginning nor the ending, but a letter representing an English vowel sound in between) was provided.

After preparing the first drafts, the tests were checked with a Persian as a first language applied linguist. The tests were also piloted with a considerable number of Persian as a first language graduate students to ensure that the items did not have potential synonyms.

Target words

Brezina and Gablasova (2013) offered an objective approach to the development of the New General Service List (new-GSL) by means of examining frequent general words across a variety of language corpora. In addition, the new-GSL's lexical units are lemmas (a headword, its inflected and reduced forms, Nation, 2013). Giving the preference to lemmas enabled them to limit the wordlist to the most frequent words with greater precision in comparison to following the word family principle. Thus, the extraction of the most frequent 1,000 words from the new-GSL seems reasonable.

To extract the first 1,000 words from the new-GSL (Brezina & Gablasova, 2013), all words with ranks more than 1,000 were removed from the list. Function words needed to be removed as well because they have little lexical meaning or may have even ambiguous meaning. Following removal of function words, the new-GSL came down to 779 words that served as the list of the most frequent words for the purpose of the current research study.

It was inferred, from Nation's (1983) and Schmitt's, Schmitt's and Clapham's (2001) experience with the number of items in the Vocabulary Levels Test, that 30 words could be statistically representative of a 1,000 word level. However, 34 words were selected for each test because at least 30 would remain even if pilot studies resulted in removing a few items.

Target words for the synonym and antonym test

For the synonym and antonym test, a group of target words with transparent (easy and clear to perceive or detect) synonyms and antonyms was needed. To serve this purpose, the following criteria were used:

- a. The target words were to have available synonyms and antonyms in online contemporary English Thesaurus dictionaries – online Thesaurus Dictionary, Merriam-Webster Thesaurus, and Collins Cobuild English Thesaurus.
- b. The possible synonyms and antonyms needed to be transparent (easy to perceive), e.g. “large” and “small” as a synonym and an antonym for “big” seem transparent. However, “bill” as an antonym for “coin” does not seem to be transparent.
- c. The major meaning sense of each target word was to be considered for synonymy and antonymy. For this purpose, the online English dictionaries were consulted to determine the major meaning sense of each target word. If the dictionaries disagreed on a major meaning, decisions were made based on the corpus data – the first 100 BNC or COCA corpus sentences for the target word were consulted to choose the most frequent meaning sense of that specific target word.

To put the criteria into effect, first, by consulting the online English Thesaurus dictionaries, the researcher (using his common sense) selected those words with possible transparent synonyms and antonyms from the 779-word list. The list was reduced to 315 words. Then, these 315 words were randomized within an Excel spreadsheet and the first 34 words were selected. However, if the meaning of the items overlapped, such as “modern” and “recent”, one of them was removed and another item with the same part of speech was selected.

Target words for the superordination and subordination test

For the superordination and subordination test, 34 target words were required which could fall into a main category (superordinate) with examples (subordinates) within the same category. For example, DOCTOR and POLICE could fall into JOB as a main category (superordinate) and WORKER and MANAGER could be examples within the same category (subordinate). For this purpose, the 779-word list was investigated to find all potential and possible relationships. An attempt was made to categorize all related words based on the aforementioned criterion. Then, 49 potential words became candidates for the test. The words (49 items) were listed in the superordination and subordination test, and the test was given to four English L1 graduate students. Their performance on each item was compared. Thirty-four items which had consistent replies to the main category section (superordination relationship) were chosen as the target words of the test.

Target words for the form and meaning tests

The selected target words for the synonym and antonym, and superordination and subordination tests also made up the target words for the two form-meaning connection tests.

Scoring the tests

Scoring of the synonym and antonym test

Considering the substantial number of available synonyms and antonyms for the target words, a norming list for scoring the synonym and antonym test was prepared to increase the objectivity of scoring the test. In order to compile the answer norms in a principled manner, two sources of information were drawn on. First, three online thesaurus dictionaries were consulted and all available synonyms and antonyms of the target words were extracted. Second, the test was given to 20 English L1 graduate students, and if two of them or more provided synonyms or antonyms which were not in the norming list, the presence of those words were checked in the BNC and COCA. On the condition that the provided words were in the corpora, they were added to the list.

A variety of synonyms and antonyms were possible for any item, but only one synonym and one antonym were required for full credit. The score of each item was either 0 (knowledge of no synonymy or antonymy), 1 (knowledge of synonymy or antonymy), or 2 (knowledge of both synonymy and antonymy). Misspellings were accepted if the intended synonym or antonym could be discerned, because the test was to measure knowledge of synonymy and antonymy, not spellings. Therefore, a second rater was not engaged to rate misspellings.

Scoring of the superordination and subordination test

When a group of 49 potential words was given to four English L1 graduate students to select the final 34 target words for the test (discussed in Target Words), the variety of different and inconsistent answers provided by them challenged the idea of preparing a norming list for the test. Therefore, instead of preparing a norming list, inspired by Webb (2005), a common-

sense approach was used to score this test. However, two raters (in addition to myself) were employed to reduce the risk of subjective ratings. The raters were English L1 Ph.D. candidates in Education. They were trained for this purpose. Inter-rater reliability was calculated to see to what extent the raters' judgment was consistent, $r(46) = .991$, $p < .001$

The raters (the employed raters and the researcher) accepted all those answers which had a reasonable superordinate and subordinate relationships with the target words. For example, for HAPPY, any answer which could be categorized as "emotion, mood, feeling, etc.", such as "sad, angry, depressed, delighted, excited", were credited. Likewise, for TREE, any answer which could be categorized as "plants" such as "flower, shrubs, weed" and even fruits' name "mango, pear" were credited. In other words, credit was given to answers as long as a reasonable connection could be considered between the target word and the answers. Thus, "garden" in addition to "plant" was also accepted as a possible superordinate for TREE, and "water" and "flower", in addition to "shrubs" and "weed", were also accepted as TREE's subordinates. In other words, participants were not expected to provide necessarily some specific answers, and all reasonable connections between the words received full credit.

The possible scores for each item ranged from 0 (knowledge of no superordinate and subordinate) up to 3 (knowledge of both superordinate and subordinates). The few encountered misspellings were accepted because the intended superordinates or subordinates could be discerned, and the study was measuring the knowledge of such relationships, not spellings.

Pilot studies

Pilot with 10 English L1 graduate students

Ten English L1 graduate students were invited to take part in the pilot study. The instructions of the tests were explained to them, and examples were reviewed. The average time they spent on the superordination and subordination, and synonym and antonym tests were 15 and 16 minutes respectively. Eighty-five point four four percent (85.44%) of their provided synonyms and antonyms were available in the extracted synonyms and antonyms from online thesaurus dictionaries. Adding one more criterion increased the percentage of the participants' acceptable answers on the test from 85.44% to 90.14%. The criterion was as follows: if two or more English L1 graduate students provided an answer (either a synonym or an antonym), the word was double-checked in thesaurus dictionaries' near synonymy or antonymy. If the word was a near synonym or antonym in one of the thesaurus dictionaries, it was added to the norming list, if not, the word was ignored.

The pilot study also showed that two items needed to be removed from the test because their meanings seemed to overlap with the meaning of two other items. In addition, one more item was also removed from the test because of its low rate of consistency with the norming list. After removing these 3 items, the test remained with 31 items and had 91.61% response rate in the pilot for accepted synonyms and antonyms for the target words.

The results of the pilot study with the superordination and subordination test showed a high performance on the test. Ninety-eight point nine two percent (98.92%) of provided superordinates and subordinates were found to be acceptable. No target word was provided with less than 90% of reasonable superordinates and subordinates. In an attempt to have the target words with more consistent superordination, four items were removed from the test.

The superordination and subordination test remained with 30 items and 98.88% of accepted answers provided by the pilot participants.

Pilot with 3 Iranian graduate students

Three Iranian graduate students in Canada were also invited to take part in the pilot study. The form-meaning tests were administered first. They spent approximately 10 minutes on the tests. The association tests were explained and administered afterwards. Care was taken to make sure they had no difficulty understanding the instructions of the tests. The tests took approximately 45 minutes. The tests, the scoring, and the procedure seemed appropriate, and the study was ready to be conducted.

Main study

Participants

The study was conducted with 40 graduate and 6 undergraduate students majoring in Chemistry, Engineering, History, Biology, and Business from 2 Iranian universities. Their first language was Persian, and English was their foreign language. They were enrolled in a private language center which was specialized in training students for academic purposes to prepare them for TOEFL and IELTS exams. The participants had taken oral and written placement tests in that language center and their speaking and writing abilities had been evaluated by language instructors in the center. All participants were registered for the advanced level and had already started the intensive courses.

The participants had learned English continuously in private language schools for at least 3 years, to a maximum of 8 years, before their current enrollment in that specific program. Additionally, they had English courses during 7 years of junior and senior high

school, 4 years of undergraduate, and 2 years of graduate studies. They aimed for an IELTS (academic module) score of 6.5 to 7.5 or more, and TOEFL score of 90 to 100 or more, because this score range usually satisfies the basic admission requirements of universities.

Their mean score on the Vocabulary Levels Test (VLT, Schmitt, Schmitt, & Clapham, 2001) at the 2nd 1,000 level was 29 (96.82%). Their mean score on a modified version (Llach & Espinosa, 2014) of the original Productive VLT (PLT, Laufer & Nation, 1999) at the 2nd 1,000 level was 26.43 (88.11%). All participants that remained in the data pool had answered correctly more than 86.66% (the criterion mastery, Schmitt et al. 2001) of the items of the VLT and PLT at the 2,000 word level.

Procedure

The form-meaning tests, the PLT, and the VLT in hard copy were administered in 10, 20, and 15 minutes respectively. Before each test, the researcher explained the test instructions to ensure that the participants were comfortable with the tasks. The participants were informed that they could spend as long as needed on the tests. The association tests were administered afterwards in approximately 40 minutes. The suggested time for each test was 30 minutes; however, the participants were informed that they could spend as long as needed on the tests. The researcher explained the instructions of the tests before the administration and ensured that the participants were comfortable with the tasks.

Results

Form and meaning tests

Table 1 shows the descriptive statistics of the participants' performance on the form-meaning tests.

Table 1. Descriptive statistics of the performance on form-meaning tests

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
SSF	46	28.00	30.00	29.74	.53	.29
SAFM	46	19.00	31.00	27.69	3.28	10.75

Note. SSFM: Superordination & Subordination Form-Meaning; SAFM: Synonym & Antonym Form-Meaning. The mean scores are from a total number of 30 and 31 items respectively.

Table 1 (the descriptive statistics) shows that participants scored 29.74/30 (99.13%) in the superordination and subordination and 27.69/31 (89.33%) in the synonym and antonym form-meaning tests. Based on Schmitt et al. (2001) criterion mastery, the participants had strong form-meaning knowledge of the target words. Schmitt et al. (2001) considered 86.66% (at least 26 items out of 30) to be the criterion for mastery of the word levels of the VLT.

Synonym and antonym test

Table 2 shows the descriptive statistics of the participants' performance on the synonym and antonym test.

Table 2. Descriptive statistics of the synonym and antonym test

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Synonym	46	12.00	30.00	23.33	4.61	21.25
Antonym	46	12.00	31.00	24.63	4.10	16.77
Total Mean	46	28.00	61.00	47.96	7.90	62.40

Note. The test has 31 items so that the maximum scores are 31 for synonyms and antonyms, and 62 for the whole test.

Table 2 (the descriptive statistics) shows that the mean score of antonyms (24.63/31) was slightly higher than synonyms (23.33/31). The table also shows the mean score of the

test was 47.96/62. Table 3 shows the descriptive statistics of the percentage of the performance on the synonym and antonym test.

Table 3. Descriptive statistics of the percentage on the synonym and antonym test

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Synonym percentage	46	19.35	48.39	37.62	7.44	55.28
Antonym percentage	46	19.35	50.00	39.73	6.61	43.63
Total mean percentage	46	45.16	98.39	77.35	12.74	162.33

Table 3 shows that on average 77.35% of the provided synonyms and antonyms were consistent with the norming list, from which synonyms had a share of 37.62%, and antonyms had a share of 39.73%. The total mean percentage (77.35%) was less than the maximum possible percentage (100%), indicating that the participants did not seem to have full knowledge of synonymy and antonymy of the items. This did not seem very impressive for advanced students at university level.

A paired samples *t*-test was run to see whether the difference between the mean scores of the synonyms and antonyms was statistically significant. The following table shows the results.

Table 4. Results of paired samples test on the mean scores of the synonyms and antonyms

	Mean	Std.Dev	Std.Error	95% Confidence Interval for Difference		t	df	sig.(2-tailed)
				Lower	Upper			
Pair 1 Syn-Ant	-1.30435	3.69305	.54451	-2.40105	-.20765	-2.395	45	.021

Table 4 shows that the difference between the mean scores (1.30) was statistically significant, $t(45) = 2.39$, $p = .02$, indicating that the participants had greater knowledge of antonyms of the target words than synonyms.

Further analysis was conducted to examine the mean number of items provided with both a synonym and an antonym (2), with a synonym or an antonym (1), and with neither a synonym nor an antonym (0). Table 5 shows the results.

Table 5. Descriptive statistics of the items provided with synonyms & antonyms consistent with the norming list

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
2, Syn & Ant	46	7.00	30.00	19.61	5.58	31.18
1, Syn or Ant	46	1.00	16.00	8.78	3.86	14.89
No Syn & Ant	46	.00	11.00	2.61	2.60	6.78

Note. 2, Syn & Ant: items are provided with both a synonym and an antonym; 1, Syn or Ant: items are provided with either a synonym or an antonym; no Syn & Ant: items are provided with no answer consistent with the norming list. Maximum mean scores are 31.

Table 5 shows that the majority of the items (19.61, approximately 2/3) were provided with synonyms and antonyms consistent with the norming list, a lower number of items (8.78) were provided with either a synonym or antonym, and a considerably lower number of items (2.61) were provided with no correct answer consistent with the norming list.

More analysis was conducted to examine the extent to which the participants were able to correctly provide the synonyms and antonyms of the items (0, 1, and 2 for 31 items). The analysis showed that on average the minimum total number of the provided synonyms and antonyms was less than 1 (.61) while the maximum was just under 2 (1.93). The analysis also showed that on average 1.55 synonyms and antonyms out of 2 were known and provided by the participants.

Superordination and subordination test

Table following table shows the descriptive statistics of each rater's evaluation of the test.

Table 6. Descriptive statistics of the performance on the superordination and subordination test evaluated by 3 raters

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Rater 1	46	63.00	90.00	83.56	6.85	46.91
Rater 2	46	63.00	90.00	84.52	6.89	47.49
Rater 3	46	63.00	90.00	84.02	6.77	45.89
Mean of Raters	46	63.00	90.00	84.03	6.78	46.03

Note. The test has 30 items, and the mean scores are from a total of 90 (30 superordinates and 60 subordinates).

Table 6 shows the mean scores of 83.56/90, 84.52/90, and 84.02/90 from the evaluations of 3 raters. The three ratings were averaged to have one single score for the purpose of reporting the participants' performance on the superordination and subordination test. The mean score of 84.03 out of 90 was close to the maximum possible score (90/90).

Pearson product-moment correlation tests were run to ensure interrater reliability. Table 7 shows the results.

Table 7. Pearson correlation between the raters' evaluations of the superordination & subordination test

		Rater-1	Rater-2	Rater-3
Rater-1	Pearson Correlation	1	.973**	.982**
	Sig. (2-tailed)		.000	.000
	N	46	46	46
Rater-2	Pearson Correlation	.973**	1	.974**
	Sig. (2-tailed)	.000		.000
	N	46	46	46
Rater-3	Pearson Correlation	.982**	.974**	1
	Sig. (2-tailed)	.000	.000	
	N	46	46	46

Table 7 shows that there were strong correlations between the raters' evaluations. In fact, there was a strong positive correlation between the first and the second rater, $r(46) = .973$, $p < .001$, the first and the third rater, $r(46) = .982$, $p < .001$, and the second and the third rater, $r(46) = .974$, $p < .001$.

Further analysis was conducted to examine the mean scores of the provided superordinates and subordinates. For this purpose, only the answers accepted as reasonable superordinates and subordinates for the items by all three raters were considered. This, in fact, helped calculate the most reliable results for the provided superordinates and subordinates, and also for the test. Table 8 shows the descriptive statistics of the performance on the test.

Table 8. Descriptive statistics of the superordinates and subordinates

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Sup score	46	18.00	30.00	27.22	3.15	9.95
Sub score	46	43.00	60.00	56.35	4.15	17.21
Total	46	62.00	90.00	83.57	6.85	46.96

Note. Sup score: superordination mean score; Sub score: subordination mean score. The test has 30 items. For each item 1 superordinate and 2 subordinates are asked to be provided so that the superordination mean score is from a total of 30, the subordination mean score is from a total of 60, and the total mean score is from 90.

Table 8 shows a mean score of 27.22 from a total of 30 (30 items, one superordinate for each) for the superordination, a mean score of 56.35 from a total of 60 (30 items, two subordinates for each) for the subordination, and a total mean score of 83.57 from a total of 90 for the test. All three mean scores were so high that it could be claimed that there was a ceiling effect; however, their total mean score (83.57) was still slightly below the maximum possible score (90/90).

Further analysis was conducted to calculate the percentages of the provided superordination and subordination so that the results on these two could be compared with each other. Table 9 shows the descriptive statistics of the percentages of the provided superordinates and subordinates.

Table 9. Descriptive statistics of the percentages of the provided superordinates and subordinates

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Sup percentage	46	60.00	100.00	90.73	10.52	110.59
Sub percentage	46	71.67	100.00	93.91	6.91	47.79

Note. Sup percentage: the mean percentage of the provided superordinates; Sub percentage: the mean percentage of the provided subordinates.

Table 9 shows a total mean percentage of 90.73 for the superordinates and 93.91 for the subordinates. It seems the participants had a better knowledge and performance on the subordination in comparison to the superordination; however, a paired sample *t*-test needs to be calculated to see whether the difference between the mean scores of the subordinates and superordinates was statistically significant. The following table shows the results of the *t*-test.

Table 10. Results of paired samples test on the mean percentages of the superordinates and subordinates

	Mean	Std.Dev	Std.Error	95% Confidence Interval for Difference		t	df	sig.(2-tailed)
				Lower	Upper			
Pair 1 Sup-Sub	-3.18804	6.95498	1.02546	-5.25342	-1.12267	-3.109	45	.003

Table 10 shows that the difference between the mean scores (3.18) was significant, $t(45) = 3.10, p = .003$, indicating that the participants had a better knowledge of the subordinates in comparison to the superordinates.

More analysis was conducted to examine the mean number of items provided with one reasonable superordinate and 2 reasonable subordinates accepted by all 3 raters. Table 11 shows the results.

Table 11. Descriptive statistics of the items provided with reasonable superordinates and subordinates accepted by all 3 raters

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
3 answers	46	16.00	30.00	25.96	3.60	12.93
2 answers	46	.00	10.00	2.61	2.37	5.62
1 answer	46	.00	5.00	.41	.86	.74
0 answer	46	.00	8.00	1.02	1.82	3.31

Note. 3 answers: 1 superordinate and 2 subordinates; 2 answers: 1 superordinate & 1 subordinate, or 2 subordinates; 1 answer: either 1 superordinate or 1 subordinate; 0 answer: neither a superordinate nor a subordinate. The mean scores are from a total of 30 items.

Table 11 shows that on average 25.96 items from a total of 30 were provided with reasonable superordinates and subordinates while this number was considerably lower for items provided with 2, 1, and 0 answer(s) – 2.61, .41, and 1.02 from 30 respectively. The results show that the majority of the items (approximately 26 out of 30) were provided with reasonable superordinates and subordinates, indicating that the participants had a strong performance on the test and that their performance was slightly lower than the maximum.

More analysis was conducted to examine the extent to which the participants were able to reasonably provide the superordinates and subordinates of the items (3, 2, 1, and 0 for 30 items). The analysis showed that the minimum total mean score of the provided answers was 2.33 out of 3 (1 superordinate and 2 subordinates) while the maximum total mean score was 3 out of 3. The analysis showed that on average 2.78 superordinates and subordinates (from a total of 3) were known and provided by the participants. This indicated that the participants had a strong performance on the test and their performance was slightly lower than the maximum (2.78/3 compared to 3/3).

Discussion

To what extent do (a group of Iranian) EAP students have productive knowledge of semantic associations at the 1,000 word frequency level?

The results showed that the percentages of correct responses were 77.33% and 93.37% on the synonym and antonym, and the superordination and subordination tests. The participants did not provide approximately one-quarter of all possible synonyms and antonyms. This may show that advanced students at the university level may not necessarily be able to produce synonyms and antonyms for all the words at 1,000 level. It should be noted that care was taken to select target words with transparent synonyms and antonyms. In contrast, the study revealed a ceiling effect (93.37%) for knowledge of superordinates and subordinates for the words at 1,000 level.

Approximately 26 out of 30 items were provided with reasonable superordinates and subordinates, indicating that the participants gave the maximum possible performance on the majority of the items. Also, the total mean number of the superordinates and subordinates produced by the participants is 2.78 out of 3 with superordinates and subordinates having approximately the same share (comparing the mean scores of 27.22 from 30 for superordinates with 56.35 from 60 for subordinates). The mean score of 84.03 out of 90 was close to the maximum score (90/90) for superordination and subordination. Knowledge of synonyms and antonyms was clearly lower than the knowledge of superordinates and subordinates.

On average, 20 out of 31 items (an approximate mean number of items) were provided with both synonyms and antonyms consistent with the norming list in which the synonyms and antonyms seemed to have the same share (comparing the mean scores of 23.33 from 31 for synonyms with 24.63 from 31 for antonyms). This is the maximum

possible performance for approximately two-thirds of the words, indicating mastery over the majority of the items.

On average, 3 out of 31 items (an approximate mean number of items) were provided with no synonym or antonym. That is, the study shows no synonymy or antonymy mastery for over 10% of the items. In other words, advanced participants at university level may have not necessarily had the synonymy or antonymy knowledge of the words at 1,000 level with transparent synonyms and antonyms. Further analysis also shows that, on average, 1.55 (out of 2) synonyms and antonyms were provided by the participants for the test.

Previous studies on associations did not show impressive results. Schmitt and Meara (1997) report that their participants could only supply around 50% of possible associations. Schmitt (1998) found that his participants performed even less than 50% on the test in the first session of his interview. Schmitt (1999) found that only 33% of his participants' associative responses matched with his norming list. However, such studies on associations had a holistic view on associations and probably tested all possible associations (collocates, coordinates, superordinates and subordinates, synonyms and antonyms) together. The main reason for this is that the aforementioned studies required participants to provide the first 3 words they could think of when they heard the target words. These studies compared the responses with the responses provided by English L1 participants (norming list), and it was not clear what type of associations were provided both by native and non-native participants. The current study, however, tested associations in a controlled way (controlled way of evaluating superordination and subordination, and synonymy and antonymy), and the findings may not necessarily be comparable with findings of previous studies.

The fact that the aforementioned studies on associations evaluated target words from lower levels of frequency than 1K (even words from 4K and 6K) should also be noted. In other words, lower performance on the associations of the earlier studies, although not necessarily comparable with the findings of the current study, may have been the result of performing on lower frequency target words. In addition, lack of knowledge or limited knowledge of words in lower levels than 1K may not necessarily make a big difference for participants because the coverage and range of such word levels are considerably less than the 1K level.

To what extent do (a group of Iranian) EAP students have productive knowledge of form and meaning at the 1,000 word frequency level?

The results show that the participants' performance on the form-meaning tests reached Schmitt et al.'s (2001) criterion of mastery as the mean percentage on the form-meaning tests were 99.13% (29.74/30) and 89.33% (27.69/31). Laufer and Nation (1999) assert that 85% to 90% performance is satisfactory for the 2,000 word level of the productive VLT. The same percentage considered by Laufer and Nation would likely be satisfactory for the first 1,000 word level as well. The main reason could be the fact that the same approximate percentage is also considered the criterion mastery of the receptive Levels Test. Schmitt et al. (2001) considered 86.66% (26 items out of 30) to be the criterion mastery of the word levels of the VLT.

To the best of my knowledge, no study has specifically measured productive form-meaning knowledge of the most frequent 1,000 words to date; however, the current results of the form-meaning tests are consistent with what Laufer and Nation (1999) and Zheng (2009) found for the 2,000 word level. The total mean scores of their university student participants

were 17 out of 18 (94.44%) and 15.61 out of 18 (86.72%) respectively at the 2,000 word level. The findings these studies show that the form-meaning of the most frequent words (either the first or the second 1,000 words or both) may often be known at university level. The mean percentage of the PLT (Llach & Espinosa, 2014) in the current study was 88.11, which reconfirms the fact that the productive form-meaning of the most frequent words, as discussed before, seems to be known by the university-level participants in this research.

Conclusions

It is worthwhile to note that higher scores in knowledge of form and meaning and lower scores in synonym and antonymy (associations) do not seem to develop equally and in tandem. In fact, the form and meaning is baseline knowledge needed in learning vocabulary and other aspects such as associations may develop incrementally (Schmitt, 2010) with repeated exposures to words (see Webb, 2007) in varied context of use and more focused instruction.

The research suggests that it is still necessary to highlight terms such as associations (including synonymy and antonymy), especially for advanced learners. Furthermore, instructional emphasis on productive outcome of language in general and productive associational vocabulary knowledge in particular need to be highlighted. It seems it is necessary to highlight that the language received from listening and readings needs to be produced in speech and writing, and for this purpose, productive aspects of vocabulary knowledge need to be emphasized.

References

- Brezina, V., & Gablasova, D. (2013). Is there a core general vocabulary? Introducing the New General List. *Applied Linguistics*, 1-23.
- Carter, R. (2012). *Vocabulary: Applied Linguistic Perspectives*. New York: Routledge.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34 (2), 213-238.
- Coxhead, A. (2011). The Academic Word List 10 years on: Research and teaching implications. *TESOL Quarterly*, 45(2), 355–62.
- Fitzpatrick, T. (2006). Habits and rabbits: word associations and the L2 lexicon. *EUROSLA Yearbook*, 6, 121- 46.
- Fitzpatrick, T., & Munby, I. (2014). Knowledge of associations. In J. Milton, & T. Fitzpatrick (Eds.), *Dimensions of vocabulary knowledge* (pp 92-105). UK: Palgrave Macmillan.
- Gilner, L., & Morales. F. (2008). Corpus-based frequency profiling: Migration to a word list based on the British National Corpus. *The Buckingham Journal of Language and Linguistics*, 1, 41–58.
- Laufer, B., & Nation, P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? In S. Foster-Cohen, & A. Nizgorodcew (Eds.), *EUROSLA yearbook* (pp. 7-28). Amsterdam and Philadelphia: John Benjamins.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399-436.
- Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: do we need both to measure vocabulary knowledge? *Language Testing*, 21(2), 202-226.
- Llach, M. P. A., & Espinosa, S. M. (2014). Knowledge of form and meaning. In J. Milton, &

- T. Fitzpatrick (Eds.), *Dimensions of vocabulary knowledge* (pp. 60-72). UK: Palgrave Macmillan.
- Meara, P. (1983). Word associations in a foreign language. *Nottingham Linguistics Circular*, 11, 29-38.
- Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K., Malmkjaer, & J., Williams (Eds.), *Performance and competence in second language acquisition* (pp. 35-53). Cambridge: Cambridge University Press.
- Milton J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, & M. Torreblanca-López (Eds.), *Further insights into non-native vocabulary teaching and learning* (pp. 83-98). Bristol: Multilingual Matters.
- Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindquist, & L. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use new perspectives on assessment and corpus analysis* (pp. 57-78). European Second Language Association.
- Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, 5, 12–25.
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. UK: Cambridge University Press.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59-82.
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd edition). UK: Cambridge University Press.
- Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10(1), 77-89.

- Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *SSLA*, 20, 17-36.
- Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning*, 48(2), 281-317.
- Schmitt, N. (1999). The relationship between TOEFL vocabulary items and meaning, association, collocation and word class knowledge. *Language Testing*, 16, 189-216.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18 (1), 55–88.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. UK: Palgrave Macmillan.
- Schmitt, N., & Schmitt, D. (2012). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484-503.
- Schoonen, R. (2010). *The development of lexical proficiency knowledge and skill*. Paper presented at the Copenhagen Symposium on Approaches to the Lexicon, Copenhagen.
- Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27(1), 33-52.
- Webb, S. (2007a). Learning word pairs and glossed sentences: The effects of a single context on vocabulary knowledge. *Language Teaching Research*, 11(1), 63-81.
- Webb, S. (2007b). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(1), 46-65.
- Webb, S. (2007c). The effects of synonymy on vocabulary learning. *Reading in a Foreign Language*, 19, 120-136.
- Webb, S. (2009). The effects of receptive and productive learning of word pairs on

vocabulary knowledge. *RELC Journal*, 30(3), 360-376.

Webb, S., & Chang, A. C. (2012). Second language vocabulary growth. *RELC Journal*, 43(1), 113 – 126.

West, M. (1953). *A general service list of English words: with semantic frequencies and a supplementary word-list for the writing of popular science and technology*. London: Longman.

Wolter, B. (2001). Comparing the L1 and L2 mental lexicon. *SSLA*, 23, 41 – 69.

Zheng, Y. (2009). Exploring Chinese EFL learners' receptive and productive vocabulary knowledge: Implications for EFL vocabulary teaching. *The Journal of Asia TEFL*, 6(1), 163-188.