A Comparative Study on Second Language Vocabulary Development: Study Abroad vs Classroom Settings

Antonio F. Jiménez Jiménez
California State University, Channel Islands

Introduction

In an imaginary Spanish language classroom a student asks: "Mrs. Smith, What is '*pelo*' in English?" and the instructor answers with a simple and straightforward "'*Pelo*' means 'hair'". The student, satisfied with the instructor's quick answer, continues the reading of the text. This type of discourse pattern is typical of most foreign language classrooms and occurs repeatedly especially at beginners' levels. However, would we agree that the student has 'acquired' this new word? What does s/he know about its frequency of use, the collocations which this word is part of, the appropriateness of its use in different social contexts, its synonyms, antonyms, derivates, metaphorical uses, idiomatic expressions, etc.? (Nation 1990: 31) In other words, what does it mean to know a lexical item?

According to Richards (1976), the nature of vocabulary learning is dense and multifaceted, and involves more than just memorizing the meaning and form of a certain word. For him, knowing a word entails:

1. Knowing the degree of probability of encountering that word in speech or print,

2. Knowing the limitations on the use of the word according to variations of function and situation,

3. Knowing the syntactic behavior associated with the word,

4. Knowing the underlying form of a word and the derivations that can be made from it,

5. Knowing the network of associations between that word and other words in the language,

6. Knowing the semantic value of the word, and

7. Knowing many of the different meanings associated with the word. (Richards 1976: 83)

Nation (1990) took the same approach and extended Richards' assumptions a step further, incorporating several other elements into a new descriptive paradigm on the nature of vocabulary knowledge that discriminated receptive and productive abilities. This categorization, as Nation points out, should be regarded as an "idealized account, rather than a realistic description of what native speakers know about most of the words in their repertoire" (1990: 32). Richards (1976) also noticed that while native speakers' grammatical competence remains relatively stable over time, vocabulary knowledge may well continue to expand in adult life. If this is the case, should instructors teach all the different components of a new word as described by Richards and Nation? Would this be effective? Which of these components are more likely to be acquired first? And last?

A search for answers to these and other questions that will be defined below inspired the foundation of the present study. The research design employed for the experiment aimed to comparatively analyze second language vocabulary development in both study abroad and classroom settings. The development of vocabulary knowledge in classroom learners was established by including students enrolled in three different (and consecutive) levels of Spanish courses and who did not have any experience abroad with the target language. Two clusters of L2 learners comprised the study abroad group: the first of these clusters consisted of students in an advanced course who had spent an academic semester (i.e., a total of five months) in a Spanish speaking country, whereas the second cluster was composed of North American TAs of Spanish who had a more prolonged experience abroad (between a year and a year and a half). The next step was to find an appropriate instrument that could assess both quantity (size) and quality (depth) of vocabulary knowledge in each of these groups.

Developing an accurate method of measuring depth of vocabulary knowledge has been a major obstacle in determining if improvement at this level actually occurs. Testing vocabulary size seems to be an easier task. A number of test formats have been used for this purpose, for example, multiple-choice tests, matching words with synonyms or definitions, supplying translations in the L1 for each L2 word, or checklist tests, in which learners simply indicate whether they know the word or not (Read 2000: 87). Recent progress in the field of computational linguistics has allowed the development of more systematic ways of

measuring vocabulary size (Kennedy 1998; Biber, Conrad and Reppen 1998). Computer software, based on a corpus of texts in the target language, has been developed to generate a word-frequency list. Researchers then choose the range of words to be object of assessment (e.g., from 500 to 10,000) and the computer randomly selects a sample of these words that can account for this range.

Studying the quality of learners' vocabulary knowledge is more complex. Henriksen (1999), in an attempt to provide a basis for a better understanding of the concept, recognized three different dimensions of the "quality of vocabulary knowledge":

a. Partial versus precise knowledge.

b. Depth of knowledge (as described above by Richards and Nation).

c. Receptive versus productive: the distinction between having some knowledge of a word and being able to use it in speech or writing.

The second source of its complexity comes from the difficulty in accurately assessing "quality" of vocabulary knowledge as conceptualized by Henriksen (1999). One common method is to through individual interviews with learners, testing how much they know about a set of target lexical items, eliciting all aspects of the word meaning that the test-taker might know (Verhallen and Schoonen 1993, Read 1989). However, this format has some negative aspects. For instance, only a very small number of target words can be covered in each interview, and adult interviewees may feel embarrassed to admit that they do not know the meaning of a certain word. To overcome these problems, innovative alternatives have been proposed in the last few years for the measurement of depth of vocabulary knowledge, such as Paribakht and Wesche's Vocabulary Knowledge Scale. This was one of the first attempts to measure quality of vocabulary knowledge in a practical way. Test-takers are presented with a five-step scale (see Figure 1)

Figure 1: The VKS elicitation scale (Paribakht and Wesche 1997: 180)

Self-report categories

I. I don't remember having seen this word before.

II. I have seen this word before, but I don't know what it means.

III. I have seen this word before, and I <u>think</u> it means _____

IV. I <u>know</u> this word. It means _____

V. I can use this word in a sentence: _____
    i. *(If you do this section, please also do Section IV.)*

and are asked to determine which step best represents their knowledge about a certain word.

As shown in Figure 1, in the first two steps test-takers do not have to demonstrate their understanding of the word. Only in the third and subsequent steps do they have to show that they indeed know the word. Although this scale represents a sensible attempt to measure depth of vocabulary knowledge, some shortcomings in its design compromise its validity. While the test is a good method to rank learners' knowledge about a single meaning of a certain word, this instrument is not able to tell us whether or not the learner knows other meanings of the word, its metaphorical uses, its collocations, etc. Secondly, the accuracy of the test in the first two steps cannot be verified, as it relies on learner's honesty. Furthermore, if the learner writes something in step three that shows that s/he does not know the meaning of the word, the researcher would be then in a position with no evidence to determine whether the learner is in step two or, otherwise, in step one. Lastly, it can only measure one word at a time, impeding the testing of a larger number of lexical items and, at the same time, making the entire testing procedure lengthy and repetitive.

Read (2000) proposes a different type of vocabulary assessment that is based on the three kinds of associations possible between target word and associate (p. 181), namely:

a. *Paradigmatic association*: the two words are synonyms, antonyms, or similar in meaning (e.g., edit-revise).

b. *Syntagmatic association*: two words that often occur together in a phrase, that is, collocates (e.g., abstract-concept).

c. *Analytic association:* the associate represents one aspect, or component of the target word (e.g., sea-water).

Read's assessment instrument consists of three parts: the target word, and two groups of four associates and detractors; some of the lexical items of the first group have a paradigmatic relationship with the target word. The second group is formed by collocates, that is, words that are usually found near the target word. In order to avoid the potential of successful guessing, the associates are unevenly distributed between the two groups. For example, for one target word, there may be three associates among the first group of lexical items and just one in the sec-

ond group. For a second target word, there may be one associate and one in the second, etc. (see Figure 2):

Figure 2: Word-Association Test (Read 2000: 184)

sudden

beautiful   quick   surprising   thirsty   change   doctor   noise   school

common

complete   light   ordinary   shared   boundary   circle   name   party

Although the expression "target word" is used in this context to refer to the stimulus word, it should be noted that with this type of instrument all nine words are "targeted" in each item. This test overcomes the problem of testing one word at a time, as happened in the previously described methods. Furthermore, it copes with deeper level of lexical knowledge, looking at both syntagmatic and paradigmatic associations between words. In summary, this word association test seems to be a fairly reliable method for measuring depth in vocabulary knowledge.

Ife, Vives Boix, and Meara (2000) used a version of this test to assess the impact of study abroad on the vocabulary development of different proficiency groups. Previous research has suggested that, in study abroad contexts, vocabulary size increases more extensively among less proficient learners than among more proficient ones (Meara 1994b, Milton and Meara 1995, Coleman 1996, Davie 1996). Ife, Vives Boix, and Meara claimed that this finding is obscured by the nature of the instruments employed, which usually measure numerical growth alone without taking into account deeper levels of vocabulary knowledge. Accordingly, they decided to employ a research instrument that would enable them to assess the learners' progress both in vocabulary size and depth of lexical knowledge. For this purpose they used a test developed by one of the authors (Vives Boix 1995) called "three-word association test" or A3VT in order to measure the acquisition of Spanish lexicon in both levels. The test, based on a 10,000-word frequency list, consisted of 120 items, each containing three words. In each of these items, two of the three words were associated in meaning, either paradigmatically, syntagmatically, or analytically (Read 2000). The third word functioned as a distractor and was identified by native speakers as not being related to the other two. The test consisted of two parts: a translation task and an association task. The translation task aimed at assessing growth of vocabulary size. For this purpose learners were asked to provide a gloss in English for all the words they knew. The second task focused on the development of depth of

109

vocabulary knowledge. For this task, participants had to cross out the word that did not have any type of association with the other two.

The researchers distributed the A3VT to 36 learners of Spanish as a foreign language prior to going to Spain as part of a study abroad program. Twenty-one of the participants were at an intermediate level and 15 at an advanced level. It is important to note here that the division between intermediate and advanced students was determined on the basis of length of prior exposure. Intermediate learners were classified as such if they had up to three years' exposure to Spanish whereas advanced learners had more than three years of experience with the language. The participants took the same test again upon their return from their period abroad. Their first hypothesis predicted that the vocabulary level of all participants would improve both in terms of the number of words known and the depth of lexical knowledge. They also hypothesized that the level of proficiency at the beginning of the study period would influence both levels of vocabulary acquisition.

The results of the pre-test were then compared with the results in the post-test and they found evidence that supported the first of their hypotheses, that is, students' performance overall improved significantly from Time 1 to Time 2 on both the association and the translation task. The second hypothesis was tested by splitting the data according to participants' proficiency level before going abroad. However, and contrary to Milton and Meara's (1995) claim that lower level learners improve more than advanced students in study abroad programs, Ife, Vives Boix, and Meara's study found significant improvement for both tests and both proficiency groups. The results also suggested a trend "towards the advanced learners making more progress in associative knowledge than in knowledge of discrete items, and towards the reverse situation among intermediate students" (pp. 72–73).

Regardless of the interesting nature of their findings, the soundness of the study was constrained by the lack of a control group: there is no evidence available to support the claim that the improvement that the participants of this study gained during their stay abroad would not have occurred if they had continued their learning experience in a classroom setting.

The present paper aims to achieve a better understanding of the process of vocabulary acquisition by examining the development of lexical knowledge in both classroom and study abroad contexts. Taking Ife, Vives Boix, and Meara's study as a starting point, this study will attempt to determine whether development in both levels of vocabulary acquisition also takes place in classroom settings. In order to carry out this developmental study on vocabulary acquisition, four variables were identified: two independent variables (classroom learning

versus study abroad learning) and two dependent variables (size and depth of vocabulary knowledge). The dependent variables were measured using a three-word association test similar to the one employed by Ife, Vives Boix, and Meara. The development of vocabulary knowledge in the classroom context (independent variable 1) was assessed by having three groups of learners of Spanish enrolled at different and consecutive academic levels. The growth of lexical knowledge in the case of study abroad learners (independent variable 2) was measured in two groups of Spanish L2 speakers who had been in study abroad programs for different lengths of time (from six to eighteen months). This group was included in the study to be compared with the results obtained in the Ife's study.

Taking all these variables into account, I hypothesized that:

1. In classroom settings, the vocabulary knowledge of the participants will improve from level to level in terms of both vocabulary size, and depth of lexical knowledge.

2. In study abroad contexts, the vocabulary knowledge of the participants will improve over time in terms of both vocabulary size, and depth of lexical knowledge.

Methodology

A total of 87 participants agreed to collaborate in the study. These participants were distributed into six groups: Groups A, B, and C were comprised by fifty-one learners who were enrolled in courses at different levels within the curricular configuration of the Spanish program of a large Eastern U.S. Research University. Due to the small number of students in intermediate courses, participants of Group A were obtained from two different sections of the same level (Span 100), a conversation class and a reading class. Having passed one of this courses was a requirement for the students enrolled in Span 200 (Group B), as was also required that students of Group C (Span 300) had Span 200 in their transcripts. None of the participants in Groups A, B, and C had experienced language contact in Spanish speaking countries. Group D consisted of eighteen students out of the 24 that were registered in an advanced Spanish course (Span 410). They were selected on the basis of their experience in study abroad programs as they all had spent one semester (five months) in Spanish speaking countries in study abroad programs run by the university. Group E was composed by 12 North American TAs (L1 English) who were teaching Spanish courses at different levels. All of them had between a year and a year and a half of experience

Table 1: Students distribution in levels and course descriptions

Group A Span 100 (n = 20)	**SubGroup A.1.** *Intermediate conversation course.* Practice in oral expression in Spanish, with emphasis on aural comprehension, idiomatic usage, and fluency. Use of journalistic materials. (Prerequisite: 3 semesters in the Basic Language Program) **SubGroup A.2.** *Intermediate reading course.* Emphasis on rapid reading comprehension. Selected readings from contemporary Hispanic literature, social sciences, current events, etc. (Prerequisite: 3 semesters in the Basic Language Program)
Group B Span 200 (n = 16)	*Intensive grammar and composition.* Designed primarily for majors and prospective majors. (Prerequisite: Intermediate grammar, conversation or reading –Span 100)
Group C Span 300 (n = 15)	*Advanced grammar and composition through reading:* Development of advanced grammar and composition skills through reading texts by native speakers and adapting their techniques for original compositions. (Prerequisite: Intensive grammar and composition –Span 200)
Group D Span 410 (n = 18)	*Spanish for business and international trade:* Students of this class were selected on the basis of their experience in study abroad programs as they all had spent one semester in Spanish speaking countries
Group E TAs of Sp (n = 12)	*North American Teaching Assistants of Spanish.* All of them had experience in Spanish speaking countries of between a year and a year and a half.
Group F Natives (n = 6)	Native Speakers of Spanish: 3 from Spain, 1 from Puerto Rico, 1 from the Dominican Republic, 1 from Mexico.

in Spanish speaking countries. Table 1 shows the distribution of the participants in the different groups as well as the departmental description of each of the classes that participated in the experiment. The study also included a cluster of 6 native speakers of Spanish that tested the assessment instrument and functioned as a control group. The participants of this group came from Spain (3), Puerto Rico (1), the Dominican Republic (1), and Mexico (1).

All participants took a three-word association test, very similar to the A3VT developed by Vives Boix (1995) and employed in the Ife, Vives Boix, and Meara's study. Learners in Ife, Vives Boix, and Meara's research had quite different language histories from those in the present study as all their participants had lived in contact with the second language in the L2 community for a period of time of at least six months whereas only two of the five in this study had had that type of experience. The assessment instrument was accordingly adjusted to fit the new participants' levels of language knowledge. The test consisted of 50 items, that is, 150 words. These items were divided into five sets, namely, ten items for each of the five groups that participated in the study (the sixth group, the native speaker group, functioned as control group and was not the main focus of the experiment). For Groups A, B, and C, the words were selected and taken directly from the glossary sections of the different textbooks that these groups used in their courses. The group of students who had been a semester abroad (Group D) paralleled quite accurately to the type of learner that participated in Ife, Vives Boix, and Meara's study so the words for this group were randomly chosen from the test that they developed based on a corpus-generated frequency list. The ten remaining items corresponded to Group E (the North American Teaching Assistants who had been in a Spanish speaking country for at least one year) and the words were selected using the Diccionario de Uso del Español de María Moliner as it is a useful reference source for collocations, synonyms, antonyms, metaphorical meanings of the words, etc. These ten items were expected to be the most challenging ones for the participants since, even though the words that were chosen are normally frequently used even in initial stages of the language acquisition process, the type of relation between the words that comprised each of these items required a deeper knowledge of the language than just the literal translation of the words. Before distributing the test to the different groups, three native speakers of Spanish and three North American Teaching Assistants of Spanish proofread and revised it, suggesting recommendations that were all included in the final version of the research instrument. The resulting test was then delivered at the end of a semester to all the groups.

The test consisted of two different tasks. For the first, an association task, learners were instructed to cross out the word in each item that did not have any type of relationship in meaning with the other two. These two lexical items could be synonyms, antonyms, collocates, metaphors, metonyms, etc. that is, they could only be related in meaning and not in form. The second assignment was the translation task, for which learners had to provide an English equivalent for each of the 150 lexical items included in the test. For the translation task, the highest score possible was 150 points, i.e., one point per word translated into English. For the purpose of the study, any of the different meanings of each word was accepted. The maximum score for the association task was 50, that is, one point per correct association established in each item. As in Ife's study, participants were given a total of 20 minutes to complete the two tasks in the order they preferred. However, it was expected that learners carried out the translation task before they decided about the association task since the translation of the terms could offer some hints about the correct solution to the association task. This prediction was confirmed by the learners in informal interviews after the completion of the test, as they reported that they almost unconsciously translated the terms before the proceeded to solve the association task. The process, they conceded, was particularly manifest in the items that were more problematic for them. Although this finding lacks scientific validity as it is based on informal reports from the learners, it can offer new avenues for research to those who aim to better characterize the role that inner speech plays in the process of language acquisition.

Analysis and Discussion

A preliminary analysis of the results to the test revealed that all participants in all groups scored higher in the translation task than in the association task. This finding was also reported in Ife, Vives Boix, and Meara's paper. The association task represents a cognitively more challenging type of activity and requires a deeper level of vocabulary expertise that goes beyond the simple translation of lexical items from one language into the other. For instance, in item 29 (cuatro—gatos—letras), 95% of the participants were able to translate all three words in English (four—cats—letters) although only 8.7% (excluding native speakers of Spanish) knew that *cuatro gatos* forms a collocation in Spanish meaning "a handful of people". The overall percentages for all groups in both tasks are shown in Figure 3.

The group in which the difference between the translation and the association task is most obvious is Group E (English TAs of Spanish). Even though their mean in the translation task was 98.45%, they scored 13.45 points less in the association task. The results in this group clearly show that even though they

Figure 3: Overall results in both the translation and association tasks (in percentages)·
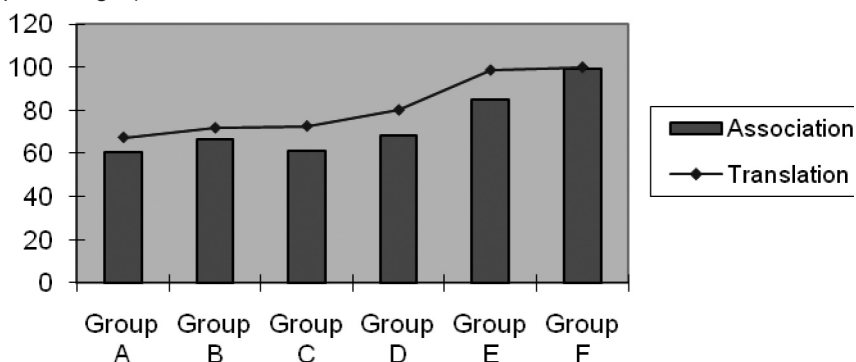
knew almost all the English words for the translation task (scoring only 1.31% less that the native speaker group), their vocabulary association knowledge is not as fully developed as the one demonstrated by native speakers (99.33%). It is in the group of native speakers (Group F) where there is almost no mismatch between the results of the two tasks.

Taking these general statements into account and in order to address our hypotheses, ANOVAs were performed on learners summed scores on both the translation and the association task. As for the translation task, an ANOVA was run to establish the effect that proficiency level (determined by each of the groups) had in the translation scores (dependent variable) producing the following results (*alpha value*=.05):

Looking at the interaction between the groups, a *post-hoc* Fisher's PLSD analysis of these results was performed (Table 3) showing no significant difference between Groups A and B (*Crit. Diff.* = 7.761; $p < 0.092$), and between Groups B and C (*Crit. Diff.* = 8.316; $p < 0.7687$). This indicates that classroom instruction did not enhance vocabulary development at the translation level. However, the standard deviation diminishes from Group A (16.165) to Group C (9.960), showing that a process of homogenization is taking place regarding learners' knowledge of vocabulary at the breadth level. On the other hand, students who spent a semester abroad (Group D) performed significantly better than these three groups (*Crit. Diff.* = 8.089; $p < 0.0072$). For their part, the North American TAs of Spanish (with at least one year of experience abroad) scored significantly higher than Group D (*Crit. Diff.* = 8.623; $p < 0.0001$). This piece of evidence suggests that while classroom instruction does not seem to have an effect on vocabulary development at the translation level, length of contact with the second language in the target community does lead to an increase in

Table 2: Means for translation

	Count	Mean Max.:150	Std. Dev.	Std. Err.
Group A	20	100.85	16.165	3.614
Group B	16	107.500	14.156	3.539
Group C	15	108.733	9.960	2.572
Group D	18	119.944	9.643	2.273
Group E	12	147.750	1.055	0.305
Group F	6	149.667	0.516	0.211

Table 3: Fisher's PLSD for Translation. Significance Level: 5%

	Mean Diff	Crit. Diff	P-Value
A, B	-6.650	7.761	.0920
A, C	-7.883	7.903	.0506
A, D	-19.094	7.517	.0001
A, E	-46.900	8.449	.0001
A, F	-48.817	10.770	.0001
B, C	-1.233	8.316	.7687
B, D	-12.444	7.950	.0025
B, E	-40.250	8.836	.0001
B, F	-42.167	11.077	.0001
C, D	-11.211	8.089	.0072
C, E	-39.017	8.961	.0001
C, F	-40.933	11.177	.0001
D, E	-27.806	8.623	.0001
D, F	-29.722	10.907	.0001
E, F	-1.917	11.569	.7425

vocabulary size. Interestingly, the group of TAs and the group of native speakers are not statistically different (*Crit. Diff.* = 11.569; *p* < .7425), which indicates that this type of vocabulary knowledge can be acquired to an extent comparable to a native's vocabulary repertoire.

However, is this type of vocabulary knowledge transferable to the association task? That is, does the acquisition of lexical associative knowledge parallel the development of vocabulary size? A separate ANOVA was run (see Table 4) to determine the interaction between the summed scores in the association task and the different proficiency groups. A *post-hoc* analysis of the results was then performed using the Fisher's PLSD method (Table 5) to establish comparative measures between groups.

As the results produced by Groups A, B, and C will require a more extended and in-depth rationalization, I will now focus on the analysis of the outcome generated by Groups D and E. These two groups, as it happened in the translation task, were significantly different according to the analysis of variance. Again, it seems that longer exposure to the target language in the second language community triggers the development of vocabulary associative knowledge. However, the Fisher's PLSD analysis reveals that there is also a significant difference between Groups E and F (*Crit. Diff.* = 4.072; *p* < 0.0008), that is, Group E did not achieve native-like competence, as was the case in the translation task. This piece of evidence shows that there are indeed two levels of vocabulary knowledge

Table 4: Means for association

	Count	Mean Max.: 50	Std. Dev.	Std. Err.
Group A	20	30.4	5.02	1.122
Group B	16	33.125	4.5	1.125
Group C	15	30.667	3.244	0.838
Group D	18	34.056	4.684	1.104
Group E	12	42.5	2.195	0.639
Group F	6	49.667	0.516	0.211

Table 5: Fisher's PLSD for Association. Significance Level: 5%

	Mean Diff	Crit. Diff	P-Value
A, B	-2.725	2.732	.0506
A, C	-.267	2.782	.8492
A, D	-3.656	2.646	.0074
A, E	-12.100	2.974	.0001
A, F	-19.267	3.791	.0001
B, C	2.458	2.927	.0986
B, D	-.931	2.798	.5101
B, E	-9.375	3.110	.0001
B, F	-16.542	3.899	.0001
C, D	-3.389	2.847	.0203
C, E	-11.833	3.154	.0001
C, F	-19.000	3.934	.0001
D, E	-8.444	3.035	.0001
D, F	-15.611	3.839	.0001
E, F	-7.167	4.072	.0008

that develop at quite different rates and ways. Vocabulary associative development requires a deeper knowledge of the lexical items that makes the process of this type of vocabulary acquisition a more slowly and complex one. Furthermore, and as the results in the association task suggest, native-like performance is harder to achieve. On the contrary, development at the translation level seems to be more easily obtainable.

As for Groups A, B, and C, the results of the association task showed no significant difference between Groups A and B (*Crit. Diff.* = 2.732; *p* < 0.506) and between Groups B and C (*Crit. Diff.* = 2.927; *p* < 0.986). However, if we look at the p-values produced by these three groups with respect to Group D, the Fisher's PLSD analysis shows that while both Groups A and C are significantly different from Group D (*Crit. Diff.* = 2.646; *p* < 0.0074 and *Crit. Diff.* = 2.847; *p* < 0.0203 respectively), Group B is not (*Crit. Diff.* = 2.798; *p* < 0.5101). Group B scored slightly higher than Group C, producing a non significant difference between Groups B and D and a significant difference between Groups C and D. The standard deviation of Groups B (4.5) and D (4.6) seemed to indicate that there were learners in Group B that performed like learners in Group D and *vice versa*. Although it might appear that this unexpected finding could be due to successful guessing, it should be noted here that the guessing factor was counterbalanced by telling learners that some invented words were inserted in the test. However, this information was actually untrue, as all words used in the test did exist. This procedure implied a certain degree of deception but it helped counteract the guessing problem and at the same time kept the translation task intact. Participants were informed by the researcher about this after the test was finished following the ethical regulations established by the Institutional Review Board of the university. Nevertheless, this procedure does not guarantee the complete neutralization of the guessing problem and does not account for these unpredicted outcomes.

The similar results produced by Groups B and D in the association task merited a closer look at the data. In order to achieve a better understanding of the situation, a separate analysis of the three different categories of associations that Read (2000) identified between the two target words was carried out, that is, paradigmatic associations (i.e., synonyms, antonyms, etc.), syntagmatic associations (i.e., collocates), and analytic associations (i.e. metaphors, metonyms, etc.). This type of lexical analysis is missing in Ife, Vives Boix, and Meara's study and in most of the literature that has looked at the acquisition of vocabulary. However, this analysis is important for it provides valuable evidence about the intricate development of vocabulary acquisition.

Table 6: Percentage of correct answers in the association task in relation to proficiency level

	Number of items	Group A	Group B	Group C	Group D	Group E	Group F
Paradigmatic	N=13	66.92%	72.07%	65.53%	69.61%	87.76%	100%
Analytic	N=20	65.75%	72.15%	69%	80.8%	92.05%	96.65%
Syntagmatic	N=17	50.29%	40.76%	45.88%	51.94%	74.47%	98.02%

The three categories of associations described by Read offer us a better interpretation of the results to the task. It can be noted that even though Groups B and D were not significantly different for the association task, they are indeed distinctive if we analyze the paradigmatic, analytic, and syntagmatic associations separately. Thus, Group D scored higher than Group B in the items where the words had either an analytic or a syntagmatic relation, but not in those with a paradigmatic association. It is also remarkable that Group A also scored higher in this type of association (i.e., paradigmatic) than Group C and only 2.69 percentage less than Group D. These results suggest that less proficient learners tend to rely more on paradigmatic communicative strategies than more advanced learners. This is to say, at initial levels of language learning, students frequently employ strategies such as circumlocutions, antonyms, synonyms, etc. to make themselves understood. On the other hand, more advanced students do not need to use them that often as they become more proficient and accurate in their language production.

Table 6 also shows that the type of association that caused the most difficulties for the learners was the syntagmatic one. This category is particularly problematic because it requires an explicit knowledge of word frequency and collocations in the target language that the other two associative categories do not need. Although the successful guessing factor was targeted by telling learners that there were invented words in the test, the problem may not have been completely deactivated. Successful guessing may have had a stronger impact on this category, as the unforeseen inconsistency among the groups shows, scoring, for instance, Group A higher than Group B, and Group C and almost equaling Group D. However, this variability in the figures disappears in Group E, where we find a quite considerable improvement (22.53 percentage of difference with respect to the preceding group). The move from Group E to F is also substantial as the figures show (from 74.47% in Group E to 98.02% in Group F).

This pattern described for the syntagmatic category does not apply to the other two types of associations. The variability in the analytic category is not that startling and, although no development is noted among Groups A, B, and C, there is an important difference in Group D with respect to the previous ones. Group E scored only 4.40% lower than the group of native speakers. For the paradigmatic category, again no appreciable difference is observed between Groups A, B, and C. As was mentioned previously, Group D, although slightly better than Group C, scored 2.4% less than Group B and only 2.7% more than Group A. There is, however, a considerable difference between these four groups and Group E and again between Groups E and F.

In summary, if we take a developmental approach in the acquisition of the three levels of associative knowledge described by Read, there is no appreciable difference between Groups A, B, and C in any of the three categories. There is a substantial difference between these groups and Group D in analytic knowledge but not in the syntagmatic and paradigmatic categories. Group E scored significantly higher in the three categories in relation to the preceding groups and almost equaling Group F in the analytic associations.

In relation to our hypotheses, the evidence collected for this study indicates that:

1. In classroom settings, the vocabulary knowledge of the language learners did not improve significantly from level to level in terms of both vocabulary size, and depth of lexical knowledge.

2. In study abroad contexts, the vocabulary knowledge of the language learners did improve over time in terms of both vocabulary size, and depth of lexical knowledge.

The second conclusion is consistent with Ife, Vives Boix, and Meara's study, as they observed that study abroad students make overall significant progress during their period abroad both in relation to the number of lexical items gained and in relation to the associative knowledge of the lexicon. However, they noticed a trend towards the advanced learners making more progress in associative knowledge that in knowledge of discrete items. This last finding is not supported in this study if we look at the overall results in both the translation and the association tasks in *Figure* 3, where the difference between Groups D and E was slightly higher in the translation task (+18.52%) than in the association task (+16.89%). In fact, the results of the present study suggest that, provided enough exposure to the target language in the second language community, native-like competence can be achieved in terms of vocabulary size, while this is harder to accomplish as regards depth of vocabulary knowledge.

Conclusions

The type of analysis conducted throughout this paper is normally absent in the vocabulary acquisition literature. The multifaceted study of lexicon development at its different levels brings to light the convoluted nature of the process of vocabulary acquisition. In order to gain a better understanding of lexicon development in all its levels and complexity, study abroad students were compared with foreign language classroom learners.

The results of the present study suggest that classroom instruction does not foster lexicon development neither in size nor in depth of vocabulary knowledge. In relation to this deeper level of lexical knowledge, there is evidence that indicates that analytic and paradigmatic associations are more likely to be acquired in classroom contexts than syntagmatic ones. In contrast, study abroad students experience an increase in both levels, improving at the breadth level to a greater extent than at the associative one. Additionally, native-like competence can be achieved in terms of vocabulary size provided enough exposure to the target language in the second language community. Conversely, native-like proficiency at the associative knowledge is harder to accomplish. A subdivision of this associative knowledge into the three mentioned categories indicates that a period of study abroad of around 6 months (Group D) seems to foster improvement in analytic associations while there is no effect in the syntagmatic and paradigmatic categories. A more extended stay in the L2 community (Group E) seems to trigger vocabulary development in all three categories. However, while there is a remarkable expansion of syntagmatic knowledge, it is nevertheless still far from being native-like.

The findings reported in this paper suggest that deeper level of vocabulary knowledge is more likely to be acquired in study abroad contexts since classroom instruction does not seem to offer the appropriate elements to trigger its development. This finding should promote future studies to look for ways to improve vocabulary teaching in classroom contexts as the methods used in today's teaching do not seem to be very effective in this aspect. This paper also aims to generate awareness among researchers in the field of applied linguistics about the importance of studying vocabulary acquisition in all its layers for it can be a good indicator of students' overall language improvement (Milton and Meara 1995). For that purpose, the tests used for its study should be further developed and improved so that they enable us to make more accurate assessments on vocabulary knowledge. The results described here should also divulge the importance that study abroad programs have in language development.

References

Biber, D., S. Conrad and R. Reppen. (1998). *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge: CUP.

Coleman, J. (1996). *Studying Languages: A Survey of British and European Students.* London: CILT.

Davie, J. (1996). Language skills, course development and the year abroad. *Language Learning Journal, 13,* 73–76.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition 21,* 303–317.

Ife, A., G. Vives Boix and P. Meara (2000). The impact of study abroad on the vocabulary develoment of different proficiency groups. *SAL, 4–1,* 55–84.

Kennedy, G. (1998). *An Introduction to Corpus Linguistics.* London: Addison Wesley Longman.

Meara, P. (1994b). The year aborad and its effects. *Language Learning Journal, 10,* 32–38.

Meara, P. (1996a). The dimensions of lexical competence. In G. Brown, K. Malmkjaer and J. Williams (eds.), *Performance and Competence in Second Language Acquisition.* Cambridge: CUP. Pp. 35–53.

Milton, J. and P. Meara (1995). How periods abroad affect vocabulary growth in a foreign language. *ITL Review of Applied Linguistics, 107–108,* 17–34.

Nation, I. S. P. (1990). *Teaching and Learning Vocabulary.* New York: Heinle and Heile.

Paribakht, T. S. and M. B. Wesche. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal 11,* 9–29.

Read, J. (1989). Towards a deeper assessment of vocabulary knowledge. *Eric Document Reproduction Service,* No. ED 301 048. Washington, DC: ERIC Clearinghouse on Languages and Linguistics.

Read, J. (2000). *Assessing Vocabulary.* Cambridge: CUP.

Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly 10,* 77–89.

Verhallen, M. and R. Schoonen. (1993) Lexical knowledge of monolingual and bilingual children. *Applied Linguistics 14,* 344–363.

Vives Boix, G. (1995). *The Development of a Measure of Lexical Organization: the Association Vocabulary Test.* University of Wales, Swansea: Unpublished PhD Thesis.