Incidental Phrasal Verb Acquisition through Second Language Reading

Amanda Hare

A Thesis

In

The Department

Of

Education
Applied Linguistics

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Arts (Applied Linguistics) at
Concordia University
Montreal, Quebec, Canada

December 2010

© Amanda Hare, 2010

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Amanda Hare

Entitled: Incidental Phrasal Verb Acquisition through Second Language Reading

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Applied Linguistics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair

Walcir Cardoso

_____ Examiner

Joanna White

_____ Examiner

Pavel Trofimovich

_____ Supervisor

Marlise Horst

Approved by _____

Chair of Department or Graduate Program Director

_____ 20____ _____

Dean of Faculty

ABSTRACT

Incidental Phrasal Verb Acquisition through Second Language Reading

Amanda Hare

Phrasal verbs are given focused attention by students, teachers and textbooks alike as being difficult to acquire but important for language learners to have some mastery over. This study examined the effectiveness of incidentally acquiring phrasal verbs and one-word verbs through second language reading with post-reading discussion activities. To my knowledge, except Bishop (2004), no other study has compared the acquisition of formulaic sequences with one-word items. The treatment involved reading eight one-page mystery stories with the targets built in for eight incidental written encounters; each story was followed by a discussion period to solve the mystery. During discussion students were able to ask for word meanings or generatively use the targets (both labeled as 'negotiation'). All discussions were recorded and later analyzed. Measures administered after treatment found eight unique textual exposures plus discussion lead to gains that were significant for phrasal verbs and one-word verbs alike. Differences in mean learning gains for the two types of verb were not statistically significant. Negotiation was able to predict both accurate and inaccurate learning; the inaccurate learning may have represented false confidence gained through negotiation; hence teachers need to be sure the correct meaning of target words are acquired through discussion. Recommendations for future research and pedagogical implications are also included. Overall, phrasal verbs may not be as difficult to learn as previously thought; in this study, they were learned as effectively as one-word verbs through eight or more incidental exposures through reading with negotiation.

ACKNOWLEDGEMENTS

I would like to acknowledge my thesis supervisor, Prof. Marlise Horst, whose time, input and guidance was invaluable in the successful completion of this thesis.

I would also like to acknowledge my thesis committee consisting of Prof. Joanna White and Prof. Pavel Trofimovich for their valuable suggestions and feedback.

I would like to thank Concordia University for giving me the opportunity to pursue this Master's degree and in turn be able to contribute to the body of Second Language Acquisition research.

Table of Contents

List of Figures Page

List of Tables Page

Incidental Phrasal Verb Acquisition through Second Language Reading

Introduction

Two personal experiences are the motivating forces behind this study on learning phrasal verbs through reading.

Why phrasal verbs? First, phrasal verbs receive special attention by students and teachers alike. It is likely that any language teacher can attest to a ubiquitous interest in phrasal verbs among students, regardless of nationality and level of English ability. In addition, textbooks, dictionaries, websites and teacher training courses have all devoted particular attention to this kind of lexical item.

Several instances have highlighted the importance of phrasal verbs during my English as a Second Language (ESL) career. Since I began tutoring during my undergraduate degree until the present as a trained professional, students have consistently expressed a burning interest in learning and mastering phrasal verbs in order to sound more natural when they speak English. During my Teaching English as a Second Language (TESL) teacher training, phrasal verbs were given special attention during the pedagogical grammar course, attention which has been reflected in almost every grammar and classroom textbook I've since consulted. The pattern of interest became undeniably apparent when my phrasal verb course, at a school for international students in Toronto, was repeatedly filled to capacity. Upon moving to Japan to teach English, the same phenomenon appeared: students were equally interested in learning and mastering phrasal verbs to make their spoken and written English appear more native-like.

The ubiquitous attention to the unassuming phrasal verb sparked my interest in doing this thesis to explore the nature of phrasal verbs and how students can better learn and master their use.

Why reading? Reading in another language has been extensively explored in second language acquisition (SLA) research, including the amount of lexis needed to understand a written text. The misunderstanding that can arise for a learner when crucial lexical items are unknown was highlighted to me in a grade ten French class. A single lexical item caused me to misunderstand a few lines of text in a short reading exercise resulting in an inability to answer the corresponding question. While it did not have a large impact on my grade, the frustration of this event has stayed with me and, in some way, has motivated my studies in vocabulary acquisition, particularly with phrasal verbs. When students encounter a new two- or three-part phrasal verb occurring adjacently, are they able to identify the verb and particle(s) as one unit? If the phrasal verb has been split by a long clause, is the student able to connect the separated verb and particle? How do students make this connection? Since phrasal verbs are always such an interest for students, what are the best ways to teach them? These are questions motivating the following thesis.

Having a strong belief in the role reading plays in increasing input of the target language and helping students to strengthen their vocabulary, I felt a natural inclination to explore whether or not reading can help learners acquire phrasal verbs and to what extent.

The following review of relevant literature will explore the role of reading, with particular focus on the acquisition of vocabulary, in SLA. First, several theories will be outlined. Then I will look at the effectiveness of incidental and intentional learning

techniques for acquiring vocabulary. This is followed by a discussion of issues in vocabulary learning through reading including: intensive vs. extensive reading, breadth vs. depth of vocabulary knowledge, and frequency findings. The final section of the literature review will examine formulaic sequences and phrasal verbs. As will be seen, the evidence presented for incidental vocabulary acquisition gives reason to believe phrasal verbs can also be learned in this manner, but little experimental evidence exists to support this notion. In addition, no studies were found exploring the role of reading-related activities on incidental phrasal verb acquisition. The goal of this research is to address these important gaps.

Following the literature review are sections outlining the methodology of an in-class experiment designed to explore the second language learners' acquisition of phrasal verbs through reading a set of short mystery stories. Details of the pilot study which contributed to planning the research are also provided.

Chapter 1: Review of Literature

Four theories relevant to incidental lexical acquisition begin the literature review. The first theory, Nagy's Default Hypothesis, is based on research involving first language (L1) vocabulary acquisition contexts. The next three theories discussed, Krashen's Input Hypothesis, Swain's Output Hypothesis, and Long's Interaction Hypothesis, deal specifically with second language (L2) learning contexts. Each is reviewed with respect to how it informs the research.

Theory 1: The Default Position

SLA studies of vocabulary learning have leaned on L1 studies as both a jumping off point and a point of comparison. It is now known that the processes used by children and adults to learn are different due to factors such as working memory capacity and background knowledge as a framework for new information. The question of how native speakers and second language speakers acquire language is very pertinent, especially applied to reading, since reading is known to be one of the best ways to acquire vocabulary, and vocabulary is the fundamental building block of any language.

Nagy, Herman and Anderson (1985) explored the amount of vocabulary native English-speaking American high school students know, and determined that it is roughly between 25,000 and 50,000 words. They surmise this lexical breadth could not have been acquired through instructed vocabulary teaching alone due to the amount of time teaching such a large number of words would require and the available class time for dedicated vocabulary teaching. These students must have incidentally acquired their large vocabularies through exposure to input in and out of school; hence the term 'default'.

The large amount of input that L1 learners receive is a key point. L1 learners are usually exposed to their native language in their home, school and, in most cases, their community environment. As they grow up they accumulate thousands of hours of passive input and active output. For L1 learners, reading in large amounts is possible since they have years to undertake this task, but for L2 learners different challenges are presented.

L2 learners face a different set of circumstances when learning their second (or third, fourth, etc.) language which points to a limitation of applying the default hypothesis to the L2 acquisition context. The hours and years of exposure L1 learners naturally have are impossible for most L2 learners to accumulate. L2 learners are generally expected to learn the vocabulary, grammar, pragmatics and possibly the orthography of the target language in a very short time, usually a few years, to a 'functional' level. Even greater difficulty is added if they are not surrounded by the target language outside their classroom, which presents a paucity of input and output practice. Written and aural texts remain a mystery until enough lexical items are known to unlock the meaning, leaving strategies such as guessing meaning from context difficult to apply. A recent study by Schmitt, Jiang and Grabe (March, 2008) confirmed earlier work reported by Nation (2006) showing learners of English need to understand at least 98% of words in a text in order to guess the meanings of the remaining unknown 2% correctly (further discussion on known word coverage occurs below). In addition, the L1 of the learner may prove to aid text comprehension or be a source of interference. The existence of similar features between two languages, such as vocabulary or grammatical structures, does not guarantee the learner will link the new knowledge to old and, in fact, similarities may be rejected (Kellerman, 2000). In some cases, interference may occur as with false

friends, e.g. the French *librarie* (a place where books are sold) and the English *library* (a place where books are borrowed). Internal or external social factors may also present barriers to learning. These factors may include: low levels of integrative or instrumental motivations for learning, inhibiting group affiliation factors and even simply lack of easy access to classes, materials and teachers. In light of and also despite these limitations, the importance of exposure to input and the role of reading are widely acknowledged in second language acquisition. The theoretical importance of input has been strongly argued in Krashen's Input Hypothesis which is the topic of the next section.

Theory 2: Krashen's Input Hypothesis

Krashen's Input Hypothesis (1985) has generated a great deal of research in the field of SLA. The Input Hypothesis posits that language is acquired incidentally by understanding messages while learning, on the other hand, is an intentional activity. An essential ingredient to language acquisition is comprehensible input (CI). The input should be at the level of 'i+1', meaning the input is slightly above the current level of the learner. In his 1991 paper, Krashen cites studies which show access to large amounts of CI through reading leads to literacy building better than other methods, such as output or error correction. To his credit, Krashen recognizes the criticism made of the Input Hypothesis that all teaching methods contain some form of comprehensible input. He counters that some contain *more* CI and these methods are superior for results (Krashen, 1991, p. 418).

As the title of Krashen's (1989) article 'We acquire vocabulary and spelling by reading: additional evidence for the Input Hypothesis' implies, he advocates that reading is a superior form of comprehensible input and more useful to language acquisition than

output through language production and Skill-Building Theory (SBT), which involves rule practice leading to automatization of the rules. The focus of reading is interpreting the message of the text, thus leaving vocabulary and spelling to be incidentally learned. Krashen admits that several methods employing some form of skill-building may show greater vocabulary acquisition than reading. However, he leans on Nagy, Herman and Anderson (1985) to make the point that reading provides frequent encounters with words in different contexts, resulting in a greater depth of word knowledge, such as collocation. He continues that this depth of knowledge cannot be transmitted as effectively through traditional instruction methods, such as direct vocabulary teaching, therefore making reading the superior method. It seems reasonable to think this principle can extend to the acquisition of phrasal verbs as well.

An example of a study conducted to test the validity of Krashen's theories of incidental vocabulary learning through reading is the research conducted by Hafiz and Tudor (1989). The researchers implemented a 12-week extensive reading program using graded readers with a group of 10 to 11 year old Pakistani students in Leeds, UK. The results showed a significant improvement in the reading and writing skills of participants on a standardized test, with the greatest improvement in writing skills, compared to a control group. While the researchers did not explicitly test vocabulary, it is tacitly understood that in order to improve one's reading and writing skills, one's vocabulary must also grow.

However, the argument that comprehensible input is sufficient for language acquisition was cast in doubt by the findings of a three year program where Francophone children in New Brunswick received English input solely in the form of books with

accompanying audio recordings with no teacher input or feedback (Lightbown, 1992). After three years, the children performed as well as students in a traditional audiolingual class. However, after three more years the students in the listening/reading program were outperformed by students in programs with teacher interaction and more written and spoken output (Lightbown, Halter, White, Horst, 2002). Therefore, contrary to the argument put forth by Krashen that reading is all a learner needs, these results show reading is useful but perhaps not sufficient. This leads us to the third theory that has relevance to the output component of the thesis reading study.

Theory 3: The Output Hypothesis

Countering Krashen's Input Hypothesis (1985), which states comprehensible input is enough for learning to occur, Swain (1985, 1995) proposes the Output Hypothesis, which presents the case that comprehensible input alone is insufficient for acquisition to occur and that comprehensible output is required for the learner's interlanguage to fully develop. Several useful functions of output are presented: fluency building, increased probability of noticing a gap in knowledge (consciousness-raising), hypothesis testing, and metalinguistic awareness, which develops when the learner uses language to discuss language as an object.

Following the idea of "pushed output", where a task forces the learner to notice problems with their L2 output, the young adolescent immersion students in Swain and Lapkin (1995) were able to successfully use their language knowledge to write a composition. A think-aloud protocol while writing and editing showed the students actively reviewed the grammar and vocabulary they used which "may have generated

linguistic knowledge that is new for the learner, or consolidated existing knowledge" (Swain & Lapkin, 1995, pg. 384).

The Output Hypothesis gives reason to argue that learners may indeed need more than encounters through reading to acquire phrasal verbs and, as such, tasks pushing learners to use the phrasal verbs may be useful. Proponents of the Interaction Hypothesis, discussed below, support this view.

Theory 4: Interaction Hypothesis

Long's (1996) Interaction Hypothesis suggests that "*negotiation for meaning*, and especially negotiation work that triggers *interactional* adjustments by the native speaker (NS) or more competent interlocutor, facilitates acquisition because it connects input, internal learner capacities, particularly selective attention, and output in productive ways" (Long, 1996, p. 451-452). The argument put forth by Long supports including interaction as a part of the classroom setting of the thesis experiment.

Learners often cite the ability to use phrasal verbs as a goal in order to sound more nativelike, but they also find it very difficult to acquire their correct usage. Long (1996) says, "Both comprehensibility and complexity are necessary for acquisition" (p. 451). According to this view, contexts involving complex structures such as phrasal verbs offer opportunities for the complex structures to become comprehensible through negotiation for meaning. Engaging with complex items on a deep level may lead to the development of a fuller hypothesis or rule for meaning and usage, and thus possibly lead to acquisition by the learner.

In summary, the above theoretical positions support the idea of reading leading to the acquisition of the meanings and uses of new lexis such as phrasal verbs. While none of the theories above directly discuss phrasal verbs, it is reasonable to assume the same processes that apply to learning through comprehending input (as advocated by the Default and Input Hypotheses) also apply to phrasal verbs. However, the discussion pointed out the limitations of these theories. Most notably, exposure to comprehensible input alone may not be sufficient for acquisition in L2 contexts. The Output Hypothesis and the Interaction Hypothesis discussed above outline the benefits of adding a spoken interaction component to activities. While reading is often a silent activity, adding spoken interaction activities related to the texts allows learners to negotiate meaning in the text, which possibly leads to deeper learning of the lexis. Having considered the theoretical perspectives, I now turn to reviewing empirical studies that support these positions. The section begins with an overview of relevant vocabulary acquisition definitions before moving on to the studies themselves.

Vocabulary Learning Through Reading: Definition of Terms

Focus will first be given to two contrasting concepts that recur in these studies, incidental and intentional acquisition. As Hulstijn (2003) points out, intentional and incidental acquisition are often associated with explicit and implicit learning and the two terms have been given many different definitions.

To begin with, intentional learning is defined as occurring when there is explicit focus on building factual, declarative knowledge through studying and memorizing grammatical structures, lexical items, etc. which the learner has consciously set out to retain. This type of instruction, often termed ‘traditional learning’, can include sentence

drills or fill in the blank exercises and often occurs where learners expect to be tested on the class content. A major criticism of intentional learning is that it is not conducive for transfer appropriate processing (TAP) to occur (Hulstijn, 2003; Lightbown, 2008). TAP occurs where the learning condition matches the conditions for use. For example, students learn a grammatical structure or vocabulary item through a dialogue that they examine and then practice, eventually moving into personalizing the dialogue in order to effectively use the items outside the controlled classroom environment. As will be shown in the discussion of Laufer (2006) below, intentional and incidental learning can be effectively paired.

How has incidental learning been defined? Hulstijn (2003) says incidental learning occurs where "the learner's attention is focused on the meaning rather than on the form of the language" (p. 349). Generally, while learners are focused on one task, such as reading and comprehending a text, they may unintentionally also learn the associated vocabulary. This does not imply learners ignore the lexical items in pursuit of understanding the meaning of the text in question since comprehending vocabulary is indispensible to comprehending the text. It is important to point out, however, that it is impossible to predict or control *what* is incidentally learned (Paribakht & Wesche, 1997). Schmidt (1990) points out that noticing and awareness are necessary for acquisition to occur; hence unfamiliar vocabulary would require attention for textual comprehension, with learning as a possible by-product of the comprehension process.

Many questions still surround the factors that contribute to incidental learning, both in and outside the class. There is agreement that it does contribute to acquisition but how it can best be used in instructional contexts is the subject of ongoing research. The

design of the thesis study here is such that explicit attention was not drawn to the target items through the reading or related activities, thereby setting up conditions for incidental learning of the targets to occur. As posited in Nagy et al. (1985), teachers and textbooks do not have the time resources to teach every aspect of vocabulary, which inevitably leaves a large amount of vocabulary acquisition to incidental contexts.

Incidental and intentional learning form two sides of instructed SLA that are mirrored in two terms describing opposing kinds of reading: extensive and intensive. Intentional learning is often paired with intensive reading where learners are "exposed to relatively short texts which are used either to exemplify specific aspects of the lexical, syntactic or discoursal system of the L2, or to provide the basis for targeted reading strategy practice" (Hafiz & Tudor, 1989, p.2). Although it is not always stated, intensive reading often includes teaching targeted features that we can assume learners are expected to retain. It is important to remember that intentional learning does not preclude incidental learning; the learner is exposed to input and acquisition may occur intentionally or incidentally. On the other hand, incidental learning is often associated with extensive reading (ER). In ER learners read large amounts of text in the L2, sometimes called 'flooding', with few or no assigned tasks to perform on the material. ER is often done for pleasure, leading the learner to focus on comprehending the narrative for meaning. Tasks are not assigned which would encourage learners to learn and retain the lexical forms (or grammatical structures, etc.) making up the text. Reading for meaning is, by definition, a prime environment for incidental learning to occur. In order to understand the application of incidental and intentional learning more completely, the discussion will address the ways these modalities have been applied in

research of another paired set of concepts: focus-on-form (FonF) and focus-on-forms (FonFs).

FonF vs. FonFs

Achieving the right balance between allowing incidental learning to occur and providing direct vocabulary instruction is a delicate challenge for teachers. Waring and Nation say "a well balanced language programme should make good use of both types of learning [incidental and intentional]. One without the other is inadequate." (2004, p. 20). Although the focus of this thesis is incidental vocabulary acquisition through reading, the argument would be incomplete without reviewing research contrasting incidental vocabulary acquisition with instructed (intensive) vocabulary learning.

Before looking at the research, it is necessary to define two primary teaching methodologies within form-focused instruction (FFI). FFI has received a great deal of research attention revolving around two constructs, focus-on-form (FonF) and focus-on-forms (FonFs), as individual approaches as well as finding the most beneficial combination. As Long (1996) outlines, FonF instruction draws learners' attention to specific, possibly problematic, language features within the context of a communicative activity; the focus in FonF is on both meaning and form as opposed to exclusively form which occurs in FonFs. FonFs refers to a more deliberate study of language that has traditionally been present in language classrooms and "involves a predominant, often exclusive, orientation to a series of isolated linguistic forms presented one after the other, as in a structural syllabus, with meaning and communication relegated to the sidelines" (Long, 1996, pg. 429).

Laufer (2006) points out the default hypothesis for FFI related to vocabulary learning is that most of it will occur through reading input and meaning-centered instruction. She summarizes research showing that this position is faulty. Laufer argues FonFs ensures noticing, which occurs before an item is acquired, although noticing still does not necessarily lead to acquisition. In FonF, noticing is possible but not guaranteed.

Laufer (2006) found an advantage for FonFs by comparing an incidental learning condition to intentional learning in two phases using 158 Israeli Grade 11 high school learners of English. The twelve target words included nouns, verbs and adjectives (but no phrasal verbs). The FonF condition involved doing a reading task with a dictionary then answering comprehension questions, whereas in the FonFs condition, learners did not read but instead received a list of English words with translations and explanations followed by exercises. A surprise test that required the learners to provide the meaning for the target words showed the FonFs group performed significantly better. The second part of the experiment on intentional acquisition tested the effect of the awareness of an upcoming test. Learners in both groups had 15 minutes to memorize target words and their meanings. Immediate and delayed post-tests showed significant gains after the second intentional learning condition. She concludes incidental acquisition can occur in FonF conditions although FonFs with a subsequent intentional memorization stage is the most beneficial.

A study by Paribakht and Wesche (1997) also found an advantage for FonFs. Adult ESL students from various backgrounds underwent two different reading instruction treatments over a three-month semester. The reading treatments compared the differences on vocabulary acquisition. The first treatment, Reading Plus (RP), consisted

of reading two texts, doing comprehension questions plus doing related vocabulary exercises, whereas in the more incidental acquisition treatment, Reading Only (RO), subjects read four texts and answered comprehension questions. The target vocabulary was nouns, verbs and discourse connectors (but again no phrasal verbs). The Vocabulary Knowledge Scale (VKS) was used to test the depth of knowledge of a word. The VKS is able to sensitively measure where incremental learning occurred by asking learners to provide a synonym or translation and then, only if they are able, to use the word in a sentence which is marked for semantic and grammatical accuracy. Results using the VKS scale showed subjects significantly gained vocabulary in both treatments; the RP treatment showed greater breadth and depth gains while the RO treatment gains were significant but "their knowledge of many of these words tended to stay at the recognition level" (Paribakht & Wesche, 1997, p. 196). Vocabulary was encountered several times over the texts although the exact number of repetitions was not reported.

These findings do not dispute the validity of incidental learning; however, they do show results are enhanced when reading is combined with more intensive, instructional methods. This leads us to consider how additional elements such as oral discussion may affect incidental vocabulary acquisition.

Vocabulary Acquisition through Oral Negotiation

Wode (1999) examined the potential of English immersion programs in Germany as venues for incidental learning. Triads of students were required to solve a hypothetical dilemma and subsequently tested on general language skills including vocabulary. While this thesis will not deal with incidental learning in immersion environments, Wode is important because his work illustrates that incidental learning can occur through oral

negotiation tasks. Wode says the discussion that occurred in the triads as students worked to solve the dilemma may have helped them to acquire some vocabulary incidentally. In the thesis study, a similar environment was created where students were grouped to solve a mystery story that contained the target vocabulary items.

Additional confirmation of the possible usefulness of oral tasks for vocabulary acquisition comes from Joe's (1998) study entitled, "What effects do text-based tasks promoting generation have on incidental vocabulary acquisition?" Her study examined how the task of retelling a story would affect the amount and depth of incidental vocabulary learning. Twelve target words were chosen for difficulty and necessity to understand the 338-word text. Three groups were formed: control, experimental and comparison. The experimental group retold the story from memory only and the comparison group had use of the text for retelling. The control group took the pre- and post-test but did not receive the treatment; instead, the control group participated in their regular class between the two test times. In both the comparison and experimental groups, learners gained new knowledge of words occurring in the texts that can be ascribed to the reading and retelling activities. Also investigating the interaction of oral negotiation tasks on vocabulary acquisition is de la Fuente's (2002) study comparing the effects of non-negotiated pre-modified input, negotiation on tasks without pushed output, and negotiation on tasks plus pushed output. De la Fuente cites several studies which found that most oral negotiation revolves around lexis. In her study, two tasks targeting 10 nouns were carried out by three groups of Spanish learners. Input in all three groups consisted of instructions to complete the tasks which contained the target items. The first group received only oral non-negotiated pre-modified input in the form of instructions.

The second group, the negotiation on tasks without pushed output condition, received input and had oral negotiation, which consisted of asking for definitions during the tasks. The third group, the negotiation on tasks plus pushed output condition, had oral negotiation for meanings in addition to receiving input but this group was also required to give instructions using the target items. While all groups acquired some level of target word comprehension, negotiation plus pushed output was most effective for productive measures and retention although it matched negotiation without pushed output on receptive measures.

The plan to include an oral negotiation task as part of the thesis was supported by this research on oral negotiation tasks to facilitate vocabulary learning. Similar to Joe's (1998) retell tasks, learners discussed short (200-500 word) mystery stories with the goal of solving the mystery. The teacher did not provide definitions and, instead, learners were directed to negotiate new meanings with each other. As discussed in Wode (1999), Joe (1998), and de la Fuente (2002) above, such oral negotiation tasks resulted in phrasal verbs being used during the discussions. This use would increase the chance of target vocabulary being acquired since it would provide more opportunities to meet the targets than reading alone. The possible effects of adding the oral element are discussed in more detail below.

Now let us turn to empirical research that has investigated the effectiveness of reading for incidental acquisition and the factors that have been shown to affect it.

Vocabulary Learning Through Reading: How Effective Is It?

Waring and Nation (2004) list several issues surrounding second language reading and incidental vocabulary acquisition which the following sections will review. The issues they point out include: 1) the relationship between reading in a second language and vocabulary acquisition, 2) the amount of words needed to be known for text comprehension, 3) the rate learners acquire new words from reading, and 4) the frequency of encounters necessary to acquire knowledge of a new word.

1. Establishing the connection between L2 reading and vocabulary acquisition

The landmark study by Saragi, Nation and Meister (1978) established a clear connection between reading and vocabulary acquisition. Their research explored L1 vocabulary acquisition using the novel *A Clockwork Orange* has sparked a number of L2 replications (e.g. Horst, Cobb & Meara, 1998; Hulstijn, 1992; Pitts, White & Krashen, 1989) as well as other focused research into L2 vocabulary acquisition through reading. In Saragi, Nation and Meister's study, adult L1 speakers read the book expecting to answer comprehension and literary criticism questions, but a few days after completing the reading, they were unexpectedly also tested on the *nadsat*, or slang words of Russian origin, from the book. Readers scored an average of 76% correct on a multiple choice test of 90 *nadsat* words (chosen from a possible 241). The methodology of following a reading passage by a surprise test has become the standard way of investigating incidental vocabulary acquisition through reading. An early replication and use of this methodology was by Pitts, White and Krashen (1989). They had two groups of adult L2 subjects read the first two chapters of *A Clockwork* Orange. Testing showed a small but significant number of the *nadsat* words were incidentally acquired, with Group 1 at 6.4%

and Group 2 at 8.1% acquisition. A critique of studies replicating Saragi, Nation and Meister (1978) appeared in Horst, Cobb and Meara (1998). They highlighted the use of short sections of text in most of the studies which resulted in a low amount of available encounters with target words. Frequent exposure to words is now understood to be important for vocabulary acquisition (frequency will be dealt with in more detail below).

In the Horst, Cobb and Meara study (1998), university students followed along while a teacher read aloud a simplified version of *The Mayor of Castorbridge* over 10 days. Students did not have access to dictionaries during reading or to the text outside of class. Contrary to the earlier replications of the *Clockwork Orange* study that had used small amounts of text, learners in this study read the entire book, which offered greater opportunities for more exposure to target items. Comparisons of the pre- and post-tests indicated that participants acquired 22% of the unknown items on a multiple choice test and a 16% gain on a word association test. These findings give stronger support to the position of L2 vocabulary acquisition through reading. Another strength of the study by Horst et al. is its longitudinal design that more closely mimics real-life reading situations where readers tend not to finish a book in one sitting (as opposed to prior replications which were conducted and tested on the same day).

Further evidence of the connection between reading and incidental vocabulary acquisition comes from Elley (1991) whose longitudinal study deals with young learners. He reported on nine similar studies that investigated the importance of comprehensible input and the effectiveness of reading for language acquisition in a year-long book flood study conducted in the ESL programs of elementary schools in Southeast Asia and the South Pacific. During the book flood students read or teachers read aloud motivating

level-appropriate storybooks after which students did extensive related activities. Although not all classes carried out the experiment identically, the researchers felt the large number of participants balanced out any differences. The results showed students taught with the traditional audiolingual system had poorer results than participants in the book flood who showed significantly superior results in the target second language (English) as well as, surprisingly, in math and native language skills. The most significant results were seen in classes where the teacher read aloud to the students. Elley credits the strong incidental language learning results to the large amounts of input through reading.

Finally, a case study by Grabe and Stoller (1997) also showed the effects of reading input on vocabulary acquisition. In this case an adult academic and total beginner of Portuguese underwent a five-month study of the language while living in Brazil. The participant read authentic written texts, mainly newspapers, and listened to some spoken texts, mainly TV news, in addition to a short three week period of instruction at the beginning of the study. Vocabulary tests showed the participant's vocabulary size substantially increased over each successive month.

In summary, these studies showed consistent findings across a variety of learning contexts, across different ages of learners, and across different languages, thus clearly confirming that incidental vocabulary acquisition can and does occur through reading in a second language. However, questions remain unanswered. For instance, none of these studies explicitly looked at the effectiveness of reading to acquire specific kinds of lexical units, such as phrasal verbs. This concern will be addressed later. The review now turns to a common question that arises for many new language learners. Namely, is reading an effective technique for beginners to learn a language and/or vocabulary?

2. The amount of words needed to be known for text comprehension

The second theme identified in Waring and Nation's overview (2004) of issues surrounding second language reading and incidental vocabulary acquisition is word coverage, or the proportion of words needed to be understood in order to comprehend a written text. A range of figures has been presented on the vocabulary breadth necessary for optimal reading comprehension or, viewed from another angle, the minimum threshold of unknown words necessary for comprehension. Laufer and Ravenhorst-Kalovski (2010) cite knowledge up to and including the 4-5,000 frequent word families based on lists compiled by Nation (2006) from the British National Corpus (BNC) as the minimum threshold for reading comprehension in English. Their findings showed learners with knowledge of these word family levels were able to understand 95% of the word meanings of texts used to judge English proficiency for entrance to university. More importantly, this level of known words represented the difference between comprehension and non-comprehension (measured in terms of responses to reading comprehension questions). They cite knowledge of 6-8,000 word families as necessary for 98% known word coverage which gives an "optimal threshold" (p.25) for reading comprehension. Earlier studies supported word knowledge coverage of the 3,000 level (Coady, Magoto, Hubbard, Graney and Mokhtari, 1993; Laufer, 1992) as necessary to reach the 95% known word threshold in a text. Nation (2006) identifies a higher 98% known word coverage criterion as necessary for reading comprehension of unsimplified texts; this corresponds to a minimum vocabulary size of 8-9,000 word families for reading and 6-7,000 for listening comprehension. Schmitt, Jiang and Grabe (March, 2008) found knowledge of 98% of words in a text being necessary to achieve 70% on a

reading comprehension measure although they stated that even "100% coverage only lead to 75% comprehension, so successful reading requires more than vocabulary, but high vocabulary levels are clearly a key requirement" (March, 2008). Hence, while 95% word coverage offers the "minimal threshold" (Laufer & Ravenhorst-Kalovski, 2010, p.26) for reading comprehension to occur as judged by reading comprehension scores, the level of known word coverage to adequately comprehend a written text is now widely accepted as 98%.

How do learners with little to no known vocabulary proceed towards achieving 98% known word coverage when they are just beginning to learn a language? If reading presents an important method to acquire vocabulary but learners do not have enough known word coverage to read well, how can they best use reading as a language learning technique, if at all? Is it better to use authentic material, which may be too difficult to comprehend as it is produced for native speakers, or is material specifically produced for language learners better, although this material may not contain enough new words to help learners quickly access more difficult texts? These questions form the 'beginner's paradox'.

Laufer (1997) reports that reading comprehension, by learners of any level but especially beginners, is hindered in three aspects: unknown words in a text, miscomprehension of words the learner thinks are familiar but in fact are unknown, and the unreliability of guessing a word's meaning from context. Being able to guess a word accurately is hindered more when the surrounding words are unknown or contextual clues are either insufficient or misunderstood (Laufer, 1997). Lack of vocabulary comprehension makes it difficult to use most reading strategies, such as guessing

unknown words from context (Bensoussan & Laufer, 1984). Grabe and Stoller say their study suggests that "learning to read in a second language centrally involves learning words" (Grabe & Stoller, 1997, p. 119). One answer to the beginner's paradox is assisting learners through instruction with intentional learning techniques to learn the thousands of basic word families which Laufer (1992, 1997) and others state are necessary to comprehend a text (Coady, 1997). Graded readers offer another solution. Beginners may use them in tandem with the learning of a basic vocabulary set or as a stand-alone tool. Graded readers are simplified extended narratives structured so that only a few words per page are unknown. They present learners with the opportunity to develop sight recognition of words and build depth and breadth of vocabulary knowledge while creating an environment for incidental learning and a pleasurable reading context.

The analysis conducted by Nation and Wang (1999) helped answer the question of how reading graded readers assists learners to incidentally build vocabulary knowledge. They analyzed 42 graded books from the Oxford Bookworms series representing seven texts from each of the six levels. They point out that while lower levels of graded readers may initially present a large vocabulary load to learners, the vocabulary learning load lessens as the learner progresses through the levels. As previously pointed out, it is impossible to begin reading without knowledge of some vocabulary so new language learners must expect to have an initially large vocabulary load to learn. In the study, a graded materials approach was used with the assumption that learners would have acquired knowledge of the 2,000 most frequent word families of English. The experimental texts were simplified to help ensure they would be easily

understood. The procedures to simplify the material are outlined in the Methodology section.

3. The rate learners acquire new words from reading

The third question about incidental vocabulary acquisition through reading identified by Waring and Nation (2004) is the rate of uptake, i.e. the amount of learning that can be expected to occur as a result of reading. Since differences in research design occur across studies (e.g. different text lengths and different testing modalities), it is difficult to say explicitly what the incidental rate of acquisition through second language reading is. Waring and Nation (2004) provide a sample of studies with different text lengths and different measurement tools whose rates of uptake range from 5.8% to 25% of target words. Waring and Nation state that while reading clearly leads to vocabulary learning, "On average, the returns are somewhat low. It seems that of the items tested about one tenth of the target words will be learned" (p. 15). However, by employing sensitive word knowledge measures, a study of extensive reading using graded readers by Horst (2005) demonstrated a group of adult English L2 learners were able to demonstrate full or partial knowledge of up to half the targeted unknown vocabulary. The measure used to test vocabulary knowledge decidedly affects at what rate acquisition is reported. For example, a word form recognition test which asks learners to respond to "Have you seen this word before?" with a yes/no answer may show a very different account of items acquired than a task testing deeper knowledge such as "Please demonstrate your ability to use this word in a grammatically and semantically correct sentence." The question of rate is closely tied to what it means to know a word.

Vocabulary knowledge has been defined using the terms breadth and depth, which are not mutually exclusive when referring to a learner's vocabulary base. A learner may have a large breadth of vocabulary and know many words by sight but may not have an in-depth grasp of the meaning(s) or uses. While the term breadth has usually been used to refer to the number of unique word families known at the recognition level where a learner can match a word form and its definition, how is depth defined?

A useful framework for understanding what depth of knowledge can include was provided by Nation (2001). He has delineated word knowledge into nine categories of knowledge, each with two layers, receptive and productive word knowledge (see Appendix A). The nine levels are organized into three global categories containing three sub-categories each. The first category deals with form consisting of spoken, written, and word part knowledge; the second, deeper, category of meaning consists of form and meaning, concept and referents, and association knowledge; while the final and deepest level of word knowledge includes use with grammatical functions, collocations, and constraints on use. A generally accepted assumption is that learning begins with making a basic form-meaning connection and then progresses on to more complex knowledge, such as collocation and use.

The breadth of a learner's vocabulary knowledge can be measured using the Vocabulary Level's Test (Nation, 2001), which has been widely utilized in vocabulary acquisition research. Many different tools have also been developed to test the depth of learner's knowledge. A well-known tool to illuminate the depth of a learner's receptive and productive knowledge of vocabulary is the "Vocabulary Knowledge Scale" (VKS). The VKS "uses a 5-point scale combining self-report and performance items to elicit self-

perceived and demonstrated knowledge of specific words in written form" (Paribakht & Wesche, 1997, p. 179). Depth is operationalized as through the ability to use a word in a semantically and grammatically correct sentence. The thesis study reported here utilized three measures to test increments of incidentally acquired vocabulary knowledge ranging from the shallowest self-rated knowledge of phrasal verb form-meaning association, through to ability to recognize simply worded English definitions correctly, and finally to deeper knowledge which involved correct grammatical use in sentence productions and recall of the particles associated with phrasal verbs.

4. The frequency of encounters necessary to acquire a new word

The fourth consideration mentioned by Waring and Nation (2004) regarding the incidental acquisition of vocabulary is frequency. A well-established connection has been made in the research literature between the frequency of encounters with a written word through reading and the learner's depth or breadth of knowledge of the item. Such research is outlined below beginning with studies dealing with the effect of frequency on mental lexical representation.

Models of the bilingual lexicon (Jiang, 2004) describe L2 words as initially sharing the same lemma, containing meaning and syntax, and lexeme, containing morphology and orthography/phonology, with the L1 translation which raises the following question: how do learners map L2 data over the L1 data, thus weakening the ties with the L1 translation and strengthening the bonds between the L2 lexical representation and the concept? One possibility is that over frequent encounters with a word a stronger form-meaning connection is developed, the mental L2 representation is more completely filled out, and knowledge of a word's use in receptive and/or productive

circumstances is deepened. If frequent encounters are necessary to create this deep knowledge, how many times does a learner need to encounter the word in order to acquire it, in the sense of being able to recognize the form, understand its meaning, draw upon the word from memory, and use it productively?

Research into the optimal frequency of encounters for vocabulary acquisition to occur has resulted in a range of numbers. Waring and Nation (2004) cite different studies with figures that range from six to 20 encounters, with equally differing vocabulary acquisition results and no definite answers on what is optimal. However, it is clear that a large number of encounters are needed to create the vocabulary depth and breadth of knowledge learners require for fluent L2 use. Proficiency can also have an effect as seen in the study by Zahar, Cobb and Spada (2001) which found the number of encounters with unknown words in incidental reading was about three to four times more important for beginners to acquire some knowledge of a word than more proficient learners who already had larger vocabularies before reading. Gains were judged by the correct amount of form to definition matches made after hearing a short textbook story, *The Golden Fleece*, read aloud then having opportunity to re-read it as many times as desired in the space of an hour. Graded readers, as mentioned above, are a controlled medium that can offer frequent encounters with new vocabulary at each level. The analysis by Nation and Wang (1999) of 42 graded readers, representing six levels of the Oxford Bookworms series, found students would need to read 5-9 texts at each level to encounter the vocabulary from each level ten times. By reading more books at higher levels, the cumulative nature of graded readers ensures the vocabulary from lower level books will also be frequently met, providing a higher possibility of being acquired.

The discussion turns to the study by Waring and Takaki (2003), which is especially insightful for the long-term durability of learning resulting from frequent incidental encounters. In their study, adult university students read the graded reader *A Little Princess* where 25 nouns and adjectives appearing between one and 18 times had been replaced by substitutions, such as *windle* for *house* and *prink* for *week*. No dictionaries or glosses were provided during the reading. Testing three months later showed that while there was a very low rate of vocabulary learning, each hour of reading produced one new word learned, and although half the words learned initially were forgotten, students better remembered words encountered more frequently.

The frequency of learner encounters with new words is also considered in Horst, Cobb and Meara's (1998) replication of Saragi, Nation and Meister (1978). Horst et al. looked at the effect of incidental textual encounters on vocabulary gains in an L2 classroom environment. The entire graded reader *The Mayor of Castorbridge* was read aloud by the teacher over six classes to two intact classes of low-intermediate ESL university students in Oman. Students followed along individually but were not allowed to use dictionaries or have access to the text outside of class. Post-tests revealed learners acquired vocabulary at a rate of one in five and "the frequency data suggest sizable learning gains can be expected to occur consistently for items that are repeated eight times or more" (Horst, Cobb & Meara, 1998, p. 215). However, as seen in the next study, different optimal rates of encounter are still being found.

A case study of a French learner by Pigada and Schmitt (2006) used a month of extensive reading of four graded readers to explore the depth of incidental vocabulary acquisition of 133 nouns and verbs resulting from frequency effects. Results showed that

partial or full knowledge of spelling, meaning, and grammatical knowledge was acquired for one out of every 1.5 words (87 out of 133 words) encountered between one and 20 or more times; twenty or more encounters were necessary for gains in all three areas of word knowledge. The frequency effects showed spelling benefitted from even a few encounters while for verb and noun meanings they found "2-19 text occurrences yielding uptake rates ranging between 16-36%" but only grammatical knowledge had "a relatively steady increase of learning along the frequency scale" (Pigada & Schmitt, 2006, p. 19).

As seen in these studies, frequency of encounter is important for building up incremental knowledge of a word across contexts, languages and proficiencies. However, no final optimal number of encounters has been arrived at in the research literature, partly due to the number of different tests used to measure knowledge and the difficulty in determining what it means to know a word. The research suggests that one or two encounters are not sufficient, yet building in the 20 repetitions recommended in Pigada and Schmitt (2006) was not possible in the mystery stories of the research. Therefore, using a rate of encounter of eight in the thesis study seemed to be a reasonable compromise. Details on operationalization of the frequency factor are given in the Methodology section.

Summary of the Incidental and Reading Literature

The above literature review has examined many different aspects of learning vocabulary through reading. Four theories were reviewed supporting varying positions beginning with the default position, which holds that a large amount of vocabulary acquisition must be done incidentally (Nagy et al., 1985), then moving to the argument that comprehensible input through reading is enough (Krashen, 1991) and then on to the

need for comprehensible output (Swain, 1985, 1995) and finally interaction (Long, 1996). Both L1 and L2 learners can profit from reading to build their vocabulary either through intentional or incidental techniques which offer more powerful learning when combined (Laufer, 2006). Other research reported above (e.g. Elley, 1991; Grabe & Stoller, 1997; Horst, 2005) add evidence to the argument for the benefit of large amounts of input for vocabulary learning as well as the positive effect of engaging extensively with texts through reading graded or authentic materials. Rates of incidental vocabulary uptake in reading experiments were shown to range widely. The research reviewed suggests that the minimum recommended coverage of known words in a text ranges from 95% (Laufer, 1992; Laufer & Ravenhorst-Kalovski, 2010) to 98% (Nation, 2006; Schmitt, Jiang & Grabe, March, 2008; Laufer & Ravenhorst-Kalovski, 2010). Frequency of encounter while reading clearly plays a role in vocabulary pick-up rates (Horst, Cobb & Meara, 1998; Pigada & Schmitt, 2006) with greater frequency of encounter being more beneficial to lower level learners (Zahar, Cobb, & Spada, 2001). The next question to be considered is how to apply these findings to the incidental acquisition of phrasal verbs through reading. The discussion will now turn to formulaic sequences and their importance before moving on to the nature of phrasal verbs themselves.

The Importance of Formulaic Sequences

Defining Formulaic Sequences

A large number of terms to describe formulaic sequences (FSs) exist throughout SLA research (Wray & Perkins, 2000). Recent years have seen an effort to pin down what formulaic sequences are and their relation to idiomaticity. Before looking at these terms, it is useful to discuss a much-quoted paper outlining the use and selection of language chunks by native speakers.

In their landmark paper, Pawley and Syder (1983) examine the ability of native speakers to do two things: 1) form grammatically correct sentences to convey meaning using natural and idiomatic utterances, and 2) maintain fluent connected discourse that is beyond the working memory capacity of the human mind. They call these two aspects "nativelike selection" and "nativelike fluency". By examining spoken texts, they argue these two features of native speech are possible by having "lexicalized sentence stems" of varying lengths stored in the mind that are drawn upon to create fixed utterances. Pawley and Syder's paper highlighted the importance of formulaticity for L1 and L2 speakers.

Broad and narrow definitions have been used to define the structure and meaning of FSs. They can range from very fixed occurrences, *above and beyond*, to more loose structures that allow for substitution, such as *the pen is mightier than the* ____, allowing for a play on words to occur (Wray & Perkins, 2000). To quote Wray and Perkins, "As Sinclair (1991) puts it: 'all evidence points to an underlying rigidity of phraseology, despite a rich superficial variation'" (p. 2). Wray and Perkins (2000) define FSs as:

> a sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar. (p. 1)

Formulaic sequences can include idioms, pre-fabricated patterns such as *in the meantime*, and common collocations like the phrasal verb *run across*. Read and Nation (2004) make a valid critique of Wray's definition by pointing out that the parameter of storage and retrieval of a whole FS may vary not only from individual to individual but also from instance to instance as factors change, such as the person's proficiency or the form of the stored item. While there may be inherent problems in Wray and Perkins' (2000) definition of FS, this definition will be used considering that the unique meaning of a phrasal verb lies in connecting the entire lexical unit of verb and particle(s) with the associated concept; if the particle is missing, the phrasal verb no longer exists. Phrasal verbs are often categorized as formulaic sequences. As the following study shows, two-word phrasal verbs are very frequent in English.

In a unique study, Gardner and Davies (2007) conducted a corpus analysis to identify the most frequent phrasal verbs in the 100-million-word BNC. Although 518, 923 phrasal verb occurrences were identified in the BNC, it was discovered that half of this large number was made up of 20 lexical verbs and eight adverbial particle combinations; in other words, one-third of all phrasal verbs occurrences are the same 25 phrasal verbs. The authors were also able to identify 100 high-frequency phrasal verbs. A possible criticism of Gardner and Davies' study is that the BNC was only analyzed for two-word phrasal verbs, while a large number of three word phrasal verbs are listed in resources such as the *Collins COBUILD Dictionary of Phrasal Verbs.* The total number of potential phrasal verbs contained in the BNC is therefore unknown. Yet this corpus research clearly shows that learners' interest in phrasal verbs as an important language feature is justified and also identifies phrasal verbs learners are more likely to know or

need to know. The design of the thesis study required targeting low-frequency phrasal verbs learners would be unlikely to know, so the list produced by Gardner and Davies was used to help exclude two-word phrasal verbs that learners would be likely to already know from the list of target items. Three word phrasal verbs were not used as frequency counts are not available.

L2 Fluency and Formulaic Sequences

The corpus-informed research discussed above shows formulaic sequences can both be identified and are frequent, but *why* are FSs important to learners? There are a number of reasons we will now turn to. Firstly, FSs have a deep cultural link which is intuitively sensed by L1 and L2 speakers. Understanding or creating new FSs requires a cultural knowledge that learners may not have; uses and new creations are accepted or rejected by native speakers based upon the feeling that the learner is trying to be part of the native speaker's culture (Prodromou, 2003). Secondly, FSs are helpful for producing fluent, rapid speech. Prodromou (2003) explains this citing the 'idiom principle' formulated by Sinclair as such: "[T]he principle of idiom is that a language user has available to him a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments' (Sinclair, 1991:110)" (p. 43). Pawley and Syder (1983) also note native speakers are able to draw on chunks of language which are combined in socially and grammatically acceptable ways at a fluent rate of speech. This chunking frees up cognitive processing to permit difficult concepts to be expressed since the mind is not required to build sentences from individual words, which is more time consuming than building sentences from larger blocks of pre-formed collocations. For example, an idea can be expressed using a 14-

word sentence created by the mind either choosing 14 individual words from memory, while also performing all the required syntactical and semantic analysis, or by choosing three pre-formed and stored chunks. Obviously, three steps are much less cognitively demanding than 14 thus leaving more working memory capacity for other mental functions. Native speakers and language learners alike benefit from such cognitive efficiency. Let us now turn to studies relating to the acquisition of FSs.

Studies of the Acquisition of Formulaic Sequences

There appear to be very few studies investigating the effects of teaching FSs to language learners. Research reported below deals with how successfully FSs are used idiomatically, possible reasons why learners have difficulty correctly using FSs, and the effect of intentionally highlighting FSs for learners as well as the effect of incidental exposure on FS acquisition. The discussion begins by looking at the function of FSs in making learners' language more idiomatic.

Acceptance, avoidance, or proper use of the second language idiom can sometimes depend on how close the idiom is to the learner's native language (Irujo, 1986a; Kellerman, 2000). Yorio (1989) refers to FSs as "conventionalized language" (p. 56) and presents research showing how FSs, or the avoidance thereof, contribute to 'idiomaticity', defined below, and the appearance of more native-like language. He points to research showing FSs can be used for a sociolinguistic, communicative function as well as a learning strategy for language development, although according to his overview of the research, FSs were not useful for grammatical development. Additionally, Yorio reports research where although the adult subjects had received formal instruction in the target language and had been operating in the second language environment, they still

showed limited control over the FSs they used, with many evident errors, which was different from findings with L1 and L2 children. For two-part verbs (phrasal verbs), he says "although the proportion of two-word verbs used by these advanced L2 learners was similar to that of native speakers, the *kind* of two-word verbs used was not the same" (Yorio, 1989, p.64). Some FSs were not employed by bilingual subjects thereby reducing the idiomaticity of their text. In contrast, a third reported study showed the texts using more FSs were subjectively considered more native-like. Yorio states:

> Idiomaticity, or native-like quality in written language, appears to be a property characterized primarily by the presence of collocations and/or sentence stems rather than by actual idioms. In second language performance, idiomaticity is further characterized by the absence of grammatical errors and by the use of quantitatively appropriate amounts of certain language-specific features, such as phrasal verbs or personal pronouns, as yet largely unidentified….although fluency is possible without grammatical accuracy, idiomaticity is not. Idiomaticity then becomes an excellent indicator of bilingual system proficiency. (p. 68)

How does a learner understand which specific words constitute a formulaic sequence in a sentence? That is, which words together form a unique meaning or a common collocation, and which combinations make grammatically correct sentences but do not constitute a common sequence? Prodromou (2003) makes an interesting point about why more advanced learners who have studied the language may still have difficulty correctly using FSs. Some of the reasons he presents for this lack of acquisition are the use of teaching materials based on non-natural English which do not adequately represent idiomatic English use (Prodromou, 2003; Irujo, 1986b) and failing to teach collocation awareness as a strategy. He also points out that while a "standard English" (p. 47) may be desirable, what makes up standard English? What FSs should be included or excluded? The ability to understand natural, idiomatic English comes from

comprehending not only the form and meaning of FSs but also the cultural nuances of the FS. Research, as discussed below, has begun to address this question; yet, more studies that approach the question from an incidental standpoint, such as the thesis research reported here, are needed to flesh out how FS acquisition occurs.

One study focusing on the difficulty of identifying FSs in a text was conducted by Bishop (2004). To investigate the relationship between the relative invisibility of FSs in texts and acquisition, Bishop looked at the effect of salient glossing of 10 FSs and 10 one-word items during a computer-based reading exercise where subjects could click on hypertexted items (enhanced or not) to get a gloss. Salient FSs were looked up more than non-salient FSs and FSs were glossed more than one-word items. Saliency on both one-word items and FSs resulted in higher scores for the experimental group on the comprehension questions which required knowledge of the targets than the control group whose on-screen words had no saliency. However, when FSs were considered alone the differences in correct answers between the treatment and control group were non-significant. This indicates that not being able to identify a FS may negatively affect acquisition. Acquisition may be further impeded with two- and three-part phrasal verbs that may have a clause or phrase separating the verb and particle(s).

A second study investigating FS acquisition, this time from an incidental standpoint, was conducted by Schmitt, Dornyei, Adolphs and Durow (2004). They investigated the change in incidentally acquired knowledge of FSs by 70 international students enrolled in a two- or three-month ESL program (this research did not consider type of instruction or frequency of exposure effects). Twenty target FSs were chosen because they generally occurred frequently in language use and occurred at least once in

the course material; an intentional learning element was introduced when instructors were asked to highlight the target FSs at least once during the course. Pre-tests showed participants had considerable receptive and productive mastery over the target FSs before beginning the course (on average 17/20 receptive and 13/20 productive). Knowledge of the target FSs increased on receptive measures by 12% (19/20 on average) and productive gains were the strongest at a 24.9%. Productive gains may appear stronger simply due to the larger learning room available. This study does not answer questions about the effects of instruction or exposure on incidental learning of FSs; however, it does support the previous argument that incidental learning of FSs can occur.

To sum up, it is evident that FSs are important vocabulary elements since they can mark one's speech as idiomatic as well as create fluent speech by allowing chunks to be strung together, helping to free up working memory for other cognitive functions. Research also shows FSs can be acquired incidentally. The next section will look at the difficulties of acquiring a particular kind of formulaic sequence, the phrasal verb.

The Difficult Nature of Phrasal Verbs

The phrasal verb (PV) as a lexical unit has often been grouped under the larger heading of FSs, yet this lexical item presents its own unique difficulties such as problems of definition, learner avoidance of use, and complexity of grammatical construction. These points will be discussed below.

As pointed out by Gardner and Davies (2007), a conclusive definition of phrasal verbs has not yet been reached. For the purpose of their research of the BNC, they chose any two-word verb+particle combination that had the particle directly adjacent or separated by one or more words. The *Collins COBUILD Dictionary of Phrasal Verbs*

(2002) considers phrasal verbs as "combinations of verbs with adverbial or prepositional particles" (p. v); verbs are considered verb+preposition/prepositional particle when it is necessary to place a noun group after the preposition although some particles can function as both adverb and preposition (p. xv). For the purposes of this thesis, a working definition of phrasal verbs will be created by combining the two definitions by Gardner and Davies (2007) and the *Collins COBUILD* Dictionary as follows: a combination of any verb plus two or more adverbial or prepositional particles that may be directly adjacent to or separated by one or more words.

Research indicates the difficult nature of phrasal verbs may lead to avoidance causing learners to choose a single word synonym instead. Avoidance occurs whether phrasal verbs are present in the L1 or not, although subjects whose L1 did not contain PVs tended to avoid them more (Dagut & Laufer, 1985; Laufer & Eliasson, 1993). In addition, Hulstijn and Marchena (1989) found intermediate learners tended to avoid PVs more than advanced learners. Three common main reasons were given for possible avoidance: "(a) L1-L2 difference, (b) L1-L2 idiomatic similarity, and (c) inherent L2 complexity" (Laufer & Eliasson, 1993, p. 45). So learner avoidance of PVs could happen where: a preference occurred for the general meaning of the one-word synonym over the specific meaning of the PV; the PV was seen as being too idiomatic, too similar to the L1, or too different from the L1; or a desire to 'play it safe' resulted in the choice of the one-word synonym. As will be discussed, the semantics and structure of the PV may also lead to avoidance.

From a semantic perspective, phrasal verbs have what Laufer (1997) calls "deceptive transparency" (p. 25); that is, the meaning cannot always be deduced by analyzing the inherent parts. For example, in the sentence "Take down the book from the shelf", *take down* is much more transparent and literal (Dagut & Laufer, 1985) in meaning than *take off* in "I'm going to take off". In the second sentence, *take off* presents two possible difficulties. One difficulty lies where the learner may only know one meaning of *take off*, such as referring to the removal of clothing or an airplane leaving the ground, but does not know the extended polysemous meaning of *leave*. The other difficulty is that the form may be completely unfamiliar to the learner making it difficult to recognize the two words together have a unique meaning.

In addition, phrasal verbs present another range of grammatical difficulties. The *Collins COBUILD* dictionary (2002) sums it up nicely:

> there are restrictions on the positions in which an adverb can be placed in relation to the object of a verb. Some particles, such as *about, over, round,* and *through* can be used as both adverbs and prepositions in particular phrasal verb combinations, although in other combinations they are restricted to one word class only, either adverb or preposition but not both. Some phrasal verbs are not normally used with pronouns as objects, others are normally only used with pronouns as objects. (p. v)

Often for each PV the learner must memorize these restrictions, or incidentally learn them. Each PV is subject to being transitive or intransitive thus possibly requiring a direct object. In addition, some PVs can be separated by a noun, pronoun, or phrase, such as *bring it back* or the much longer *take the library book that is two weeks overdue back to the library.* Students must memorize which PVs can be separated and which cannot.

Passive forms are also possible for some, such as *my time was eaten up by the computer crash*.

Other difficulties lie in collocation, generation and use. Phrasal verbs can have strong collocations, such as *break up with SOMEONE* and not *break up with SOMETHING*, which must be memorized as part of the form. The frequency lists provided by Gardner and Davies (2007) offer guidance to L2 English learners as to which PVs are important to know. A second problem relates to newly generated PVs which, when subject to nativelike selection (Pawley & Syder, 1983), may or may not be accepted. However, the adverbial and prepositional particles of phrasal verbs have particular meanings and are often the basis for new formations which is helpful to the language learner in both creating and comprehending PVs. The pragmatic knowledge of when to use phrasal verbs lends yet another challenge: PV use tends to occur in informal, spoken settings although some are completely acceptable in formal settings; however, it depends on the formality of the situation and again brings Pawley and Syder's (1983) nativelike selection and use into play. All of these difficulties can be daunting for learners to overcome in the process of incorporating this new type of lexical item to their productive vocabulary. Exposure to PVs may help deepen knowledge leading to a reduction in avoidance and a more confident use. This thesis study investigated whether incidental exposure to PVs through reading can help incrementally increase knowledge of the target PVs.

Incidental Acquisition of Phrasal Verbs through Reading

The survey of research above has examined both incidental and intentional learning contexts and showed that reading does contribute to vocabulary acquisition. The research highlights the importance of frequently meeting new lexis in comprehensible texts. The usefulness of learning lexis through oral interaction was also outlined although more experimental support for this theoretical position is needed. At the time of writing, no studies have been found that specifically target the acquisition of phrasal verbs in an incidental learning context. Phrasal verbs have been included as targets in reading studies (Bishop, 2004), but were grouped with other kinds of formulaic sequences and idioms. Other studies have looked at the difficulty of PV use and avoidance related to the existence, or not, of PVs in the first language (Hulstijn & Marchena, 1989; Laufer & Eliasson, 1993). A dedicated study will allow for considerations in the testing and treatment to be sensitive to the unique nature of phrasal verbs. This thesis study investigated the effects of learning PVs with a reading task followed by oral discussion. This study is unique in that the longitudinal design and in-class component of the methodology, with discussion and related activities, more closely mimics a real L2 classroom reading situation. Thus results are indicative of what teachers and learners can expect in the everyday classroom. While previous research of incidentally acquired vocabulary has followed a read and immediate test design, the format of this study used a different model by including post-reading activities that occur before tests of PV knowledge are administered, which may cause additional uncontrolled exposures to the targets beyond those available in the experimental reading passages. In other words, this sequence, which is closer to the actual exposure to the targets learners would get in the

classroom, means that only the text input exposure will be controlled but additional activity-related exposure cannot be. As outlined in the methodology section, variables such as time on task, written instructions, and input from the teacher speaking, were carefully controlled. The three research questions being investigated in this study follow:

1. Will eight repetitions of phrasal verbs in written text be sufficient for incidental acquisition to occur (recognizing that learners will have additional exposures to the target words during the follow-up activities)?
2. Will depth of knowledge for unknown phrasal verbs and unknown one-word verbs be equal after eight contextual exposures through written text and comprehension-focused follow-up activities?
3. Are targets that are explained during communicative interaction acquired more deeply than words that are not given this kind of attention?

The following hypotheses were proposed:

Hypothesis 1: Previous research shows an exposure frequency of eight repetitions of a lexical item within written text may lead to incidental acquisition (Horst, Cobb & Meara, 1998). It was predicted eight exposures to phrasal verbs in an incidental learning setting that also included possible exposures to the words during oral interaction tasks would be sufficient for a form-meaning connection to be made leading to surface recognition of phrasal verbs as represented by a change in YES items on the self-rated test.

Hypothesis 2: It was expected that unknown single word items would be acquired more deeply than phrasal verbs although unknown phrasal verbs would show some level of learning. Difficulty of acquisition was expected due to the more complex two-word

structure of phrasal verbs which required more processing than the one-word verbs. To my knowledge only Bishop (2004) has compared the ease of acquisition of multi-word formulaic sequences with one-word items although he used saliency to aid acquisition; hence, this investigation sheds light on the unanswered question of incidental acquisition of multi-word formulaic sequences compared to one-word items.

Hypothesis 3: Target items learners asked about were expected to be acquired more deeply due to the fact that these items directly came into the learner's awareness as being unknown. As Bishop (2004) showed, saliency can lead to deeper acquisition. In Bishop's case, the learner's awareness that the target item's meaning was unknown made it salient to the learner. For this thesis study, explanations given by other learners, as the teacher directed any questions about meaning to other participants, would also add salience. Research by de la Fuente (2002) showed that acquisition of negotiated items was greater than non-negotiated items.

Chapter 2: Method

Purpose

The purpose of this experiment was to investigate the effectiveness of reading activities on the incidental acquisition of eight phrasal verbs (PVs) compared to a set of eight low-frequency one-word verbs. The small but important lexical category of PVs was chosen because it has not been investigated widely. As noted in the literature review, previous research reported above which looked at incidental acquisition has tended to focus on nouns, verbs and adjectives, and almost always on individual lexical items.

Pilot Study

To investigate the feasibility of the thesis experiment, a pilot study was carried out as part of the teaching curriculum in a regular communicative classroom. The positive learning results indicated a more controlled experiment would be worthwhile. Methodological strengths and weaknesses also became evident through the pilot study; these informed the design of the thesis experiment. For these reasons, a brief overview of the pilot study is presented below.

Research Questions and Methodology. The research questions investigated in the pilot study were similar to those outlined at the end of Chapter 1. The pilot study took place in an intact class of seven high beginner/intermediate ESL learners of various L1 backgrounds in a communicative classroom environment in a Montreal community centre. As part of a communicative general English class, the students read eight short mystery stories for homework which had been adapted to include 10 target phrasal verbs and nine one-word items resulting in eight unique contextual encounters after all eight stories were read. The goal was to determine the extent each type of item was acquired as measured by performance measures administered after the learners had read the stories.

The read-and-test design has been previously used in incidental vocabulary studies (see Chapter 1) and is also used in the reported thesis study.

Target Items. Target items were selected as likely to be unfamiliar to learners based upon native intuition and experience. Some of the targets were chosen due to their natural occurrence in the stories. The experimental targets included 10 phrasal verbs and nine one-word targets comprising nouns, verbs and adjectives that were built into the short mystery stories. Examples of the target phrasal verbs included *break up with, gulp down,* and *count on*. Lexical frequency profiling of the stories conducted using software available at www.lextutor.ca identified nine infrequent one-word targets including nouns, verbs and adjectives. Examples of one-word items are *cabin, husky* and *assert.* Another determining factor for choosing the targets was the ease of inserting the words into the stories for a frequency count of eight. In addition, 10 single word and phrasal verb control items naturally occurring once only across the eight stories were used. A total of 29 lexical items were tested in the pilot study. The pilot study compared phrasal verb acquisition to verbs, nouns and adjectives. To increase validity in the thesis study, items of the same word class were compared: phrasal verbs and one-word verbs. No other parts of speech were tested.

Issues with the pilot study target words were identified through observation or discussion with the students after the post-testing. One of the target words, *alibi*, was a French cognate that was not identified until a participant pointed it out during the post-testing. Some of the phrasal verbs were more common than the researcher had imagined, such as *break up with*. Careful selection of the target words is necessary to ensure they are low frequency items and are not cognate with French.

Reading Treatment. The eight stories were assigned as homework reading over eight classes, one per class, within the context of a larger communicative class. Students were asked to attempt to solve the mystery but not to consult a dictionary while reading. Each class began by the students discussing the story and solution. If questions arose about vocabulary, the teacher (and researcher) avoided explaining the words and directed students to discuss the meanings with each other. Pilot study students were allowed to keep all paper copies of the stories which meant they could re-read them; this design flaw raises questions about experimental validity since time on task and exposure to items were not strictly controlled. This problem was avoided in the thesis experiment by not allowing students to retain copies of the stories.

Directing students to discuss unknown meanings with each other was an effective strategy. The teacher observed students examining the context of the sentence and consulting with their partners or other students to often successfully determine the unknown word's meaning. Students primarily asked questions about target items. This showed the target words were sufficiently difficult and that the written medium of delivery made unknown targets salient. Negotiation helped enable students to correctly deduce meanings. The same strategy was used in the thesis experiment.

The delivery of the stories was changed in the thesis study. Instead of one homework reading per class, participants read two stories during class time and, as mentioned, they were not allowed to take a copy of the story home. Unrelated communicative activities were conducted as a break between the stories for the following reasons. Most importantly, an unrelated activity helps remove the target vocabulary from each student's working memory between treatments. From a pedagogical perspective, the

change of activity makes the class more interesting for students and eases the potentially heavy cognitive load of solving a mystery story in a non-native language. Instead of a communicative based general English class, the reading was the main focus of the class and time on task was held consistent across each activity. The restriction on dictionary use remained to help ensure the source of learning was solely the exposure in the texts.

Quantitative Tests. The following tests were used as pre- and post-test measures of vocabulary knowledge in the pilot experiment. The pilot test quantitative measures generally showed positive learning gains. Comprehensive results of the pilot study tests are not reported since these measures were altered or omitted from the thesis study for reasons outlined below.

1. Adapted version of Nation's Levels Test (2001): Ten questions were posed using the same layout as the Levels Test of presenting six unknown words to be matched with three possible definitions. See sample item below. There were a total of 60 one-word items and phrasal verbs presented. The 29 target items were randomly scattered throughout. The definitions were either correct answers or unrelated distracters. The test effectiveness was lowered by an oversight wherein the same set of distracters was used repetitively through the test questions. The Levels test question format was used to test receptive knowledge of the target items in the pilot study and in the thesis study reported below.

   A sample item is as follows:

   1 abruptly
   2 clean out __ take all the money and valuables
   3 fuming __ to be extremely, but quietly angry
   4 ransom __ change your position to standing, not sitting or lying
   5 scowl
   6 stand up

2. Short answer test: Students were asked to answer 29 short answer questions using a provided word. The particle(s) of the phrasal verbs were not provided to help determine if students had acquired both the meaning of the phrasal verb and the associated particle. This appeared to be an effective test of deep acquisition of collocation as well as productive knowledge of phrasal verb form. This test was used again in the thesis study but only in post- and delayed post-tests. Here are some examples highlighting the difference between the pre- and post-test answers from the pilot study:

    Target PV: clean out
    Test question: Michelle knows a family who was robbed on Christmas Day. Do you know anyone who was robbed? (clean)
    Pre-test: Yes, I do, in Brazil almost all my friends was robbed. I know the case which the house was totally clean.
    Post-test: Yes, the robber clean out my friend's house.

    Target PV: turn up
    Test question: Some people don't like surprises. What about you? Do you like people to phone before they visit? (turn)
    Pre-test: Yes, I don't like people that turn over* without telling me a word.
    Post-test: I really hate when people turn up without calling me before.
    *Note that the student correctly provided a particle although the intended phrasal verb was *turn up*.

3. Word order test: Students had to put ten jumbled sentences, testing only phrasal verbs, in the correct order. This test helped show whether the participants could correctly identify the phrasal verb particle(s) as part of the verb and if the particle(s) could be correctly placed in the sentence. Several students found this test extremely difficult and it took a long time. This test was not used again as the

difficulty of this type of test and time constraints unduly strained participants.

Here are some example answers provided by students during the pilot study:

*I spent out at home, hanging all weekend.
*The audience was boring so the concert walked out.
*I will try to spell out so you don't understand it.
*A thief cleaned out my entered apartment and me.

Limitations. Several limitations were identified within the methodology of the pilot study. As described above, the quantitative measures were found to be either poorly constructed (as in the case of overusing the same distracters in the adapted Levels Test), were too long (as in the short answer test), or were too difficult (as in the jumbled sentence exercise). The two main issues which arose from not pre-testing the target items, namely limited room to demonstrate growth due to the presence of high-frequency items or the inclusion of cognates, are easily avoided by more careful target word selection.

Finally, the choice to assign the reading exercises as homework led to several issues corrected in the study related below. It was observed as the pilot study continued that several students were not completing the homework reading assignments. At the beginning of class some students would sit in silence reading while their partners waited to discuss the solution. However, this was informative as I could determine how long it takes a student to silently read a story. By allowing the students to keep a copy of the stories, the possibility existed for students to compare the stories or to re-read the texts outside the class, thus threatening the experimental control over frequency of exposure of eight times. In addition, since the students read the stories at home they may have used dictionaries to look up unknown or partially known words. The thesis study below corrected these limitations by only allowing access to paper copies of the stories during class time without dictionary use. To ensure all students received equal exposure, the

stories were read out loud once by the teacher and then equal time on task was allotted to solve each of the mystery stories.

Thesis Study Methodology

Setting

The experiment was conducted in a classroom-based setting in an Ontario LINC (Landed Immigrants and Newcomers) centre. The main focus of the class was reading.

Learners were told the experiment was to investigate the effect of reading on language ability but they were not aware that lexis was the focus of the research.

Participants

Participants consisted of one intact group of 16 intermediate and advanced adult ESL learners gathered from several classes at the LINC centre. Students came from diverse language backgrounds, such as Russian, Filipino, Urdu, French, Bulgarian, and Amharic. They had been in Canada for an average of two years. Students had studied English for an average of 3.2 years. Although 16 students participated in the learning activities, only 14 students completed all of the experimental measures.

The majority of students were LINC level 5 (Canadian Language Benchmarks (CLB) 5, 6 or intermediate), two students were LINC level 6 (CLB 6, 7 or upper intermediate) and two students were LINC level 7 (CLB 7, 8 or advanced). Pre-testing was carried out to ensure learners did not know the target vocabulary items used in the study. Details of the pre-tests are outlined below.

The intermediate group presented greater potential in terms of 'learning room' as this population typically has a greater amount of unknown vocabulary than advanced learners. Using only the intermediate learners to achieve an ideal group of the same level would have resulted in a substantially reduced sample size. Including the upper intermediate and

advanced learners to increase the sample size is a recognized limitation of this study. Statistical adjustments were not felt necessary for the advanced level learners as they represented only two of the 14 students included in the final analysis.

Design. This comparison study uses a pre-/post-test design. As described above, the experimental treatment entailed reading eight stories with embedded learning targets followed by pair then class discussion to solve the mystery. Students were randomly placed with a new partner each class as much as possible. Pairs had access to the texts during the follow-up activities, which was more ecologically valid as an authentic classroom reading task. Although the additional exposures during the follow-up tasks were uncontrolled, the presentation of the targets in eight unique written contexts was maintained. A baseline control group was not included since it has been established by the research reported above that exposure to input leads to acquisition. Hence, comparison to a group that has not been exposed to the input would not provide new research insights. This study compared the effects of written exposure coupled with oral negotiation on acquisition of two different kinds of lexical items.

Materials

Selection of Target Words. Three criteria were used in selecting eight PVs and eight one-word target items: low frequency in English, opaqueness, and suitability to story content. Pre-tests were used to ensure the target items were unknown to participants (described below).

In addition to the PV frequency list in the article by Gardner and Davies (2007), another resource for unearthing the frequency of PVs came from the *Collins COBUILD Dictionary of Phrasal Verbs* (2002). This dictionary outlines the polysemous meanings of

48 particles showing detailed nuances of meaning. The associated headwords commonly collocating for each particle are also listed. A frequency of occurrence number for each particle is provided for the number of times it occurs in the dictionary. Both lists were used in selecting suitable, low frequency PVs for the study. The second criterion, opaqueness of meaning, was operationalized as follows. According to the *Collins COBUILD Dictionary of Phrasal Verbs* (2002), PVs have four main combinations (p. vi) which are:

- Non-literal where the meaning is not easily understood by individual parts or the combination of the parts, for example, *go off* = explode
- Fixed particle where the verb and preposition or adverb are common collocations or commonly occur together for a particular meaning, for example, *lead to*
- Completive-intensives where "the particle does not change the meaning of the verb, but is used to suggest that the action described by the verb is performed thoroughly, completely, or continuously" (p. vi), for example, *slave away*
- Literal or semi-literal where the meaning is easily understood and the verb and particle(s) commonly collocate together although both verb and particle(s) can be found in other combinations with different meanings, for example, *fight back* and *hit back*

Non-literal phrasal verbs, as judged by the researcher's native intuition, were chosen since it was hypothesized their opaque nature would make it necessary to guess the meanings from context. Finally, as learned from the pilot study, efforts were made to keep as many naturally occurring words in the stories as possible among the total set of 16 target items.

The one-word target items chosen for the pre-test were low-frequency verbs occurring within the 9,000 level of most frequent English words; this list is one of 15 frequency lists compiled by Nation (2006) based upon the British National Corpus (BNC). Care was taken to exclude Greco-Latin cognates.

As previously mentioned, there are varying views on the amount of known word coverage needed in a text before comprehension is impeded; Laufer and Ravenhorst-Kalovski's research (2010) identified 95% as the minimum, Nation (2006) places the limit at 98%, while Schmitt, Jiang & Grabe (March, 2008) also cited 98% as the lowest limit of known words in a text. Laufer and Ravenhorst-Kalovski cite 98% as the optimal reading level. The mystery stories averaged 367 words in length. Therefore, 16 unknown words represent 4.4% of 367 possible unknown words per story which falls within the acceptable range of 5% unknown items (making 95% of text coverage as known words).

Targets were pre-selected as outlined above and written into the eight target texts. Each item appeared once in each story resulting in eight unique exposures across the stories.

Selection of Experimental Texts. The same original set of experimental texts used in the pilot study was used for the thesis study although adapted as outlined below. The texts consisted of eight short one-page mystery stories. These stories were adapted from the book *Two Minute Mysteries* (Sobol, 1967). These texts were considered ideal as the original intended audience was native-English speaking elementary school children, hence it was assumed the level of vocabulary, the grammatical constructions, and the difficulty of the mystery to solve would be at a level appropriate to the English language ability of this study's participants. In addition, presenting eight unique stories ensured the target items were met in eight unique contexts which more closely mimics natural reading. Re-reading the same text eight times would give identical contextual encounters (Pigada & Schmitt, 2006) and is a technique used more for instructed learning as opposed to learning in the natural reading environment this thesis study aimed to create.

Mystery stories were chosen in order to give learners a purpose when reading the text as well as keep their focus on overall text comprehension of the plot rather than particular words. While it is noted that linguistic clues are necessary to solve the mystery and this may result in some focused attention on the meaning of individual lexical items, no special attention was drawn to any lexical items. Instead, solving the mystery was intended as the overall motivating task with a small reward of a candy to any participant who found the solution.

The original texts were kept as authentic as possible. To increase the ease of ESL reading, changes such as speaker identification or setting change markers were added. The stories were analyzed using the lexical frequency profiling program available on the Lextutor website (www.lextutor.ca) in order to ensure that all context words, except the targets, were among the 2,000 most frequent words of English according to lists based on the BNC (Nation, 2006). Lexical items outside the 2,000 most frequent were replaced with substitutions selected by native intuition with the assistance of an ESL dictionary. Dependant clauses were added explaining necessary content words above the 2,000 level. Minor changes were made, only where necessary, to adjust the grammatical difficulty to match a high-beginner/low-intermediate level of English. All changes were judged as acceptable by two experienced ESL teachers.

Adding new sentences or rewriting existing sentences was necessary to add in the eight repetitions of the target words necessary to meet the frequency of exposure criterion. These additions were written carefully to match the voice of the original author, grammatical difficulty, and tense. The texts were read by at least one native English

speaker to judge the success of adding in the new sentences. 'Success' meant that the judge felt the text flowed smoothly and no sentences were unusual or out of place.

An excerpt of the original text and modified text are below. This sample reflects how the original text was changed to add in redundancy for ease of understanding, speaker identification, one expansion sentence pointing to the meaning of the target *bicker*, and the addition of target vocabulary.

Sample of the original text from the story "The case of the missing button".

"She isn't positive-she saw only your back. But this missing button proves you did it. Luckily, Miss Casey isn't badly hurt. Now, where's her purse?"

"Matty kept insisting he didn't know a thing about the slugging and theft," the inspector told Dr. Haledjian later.

"No doubt," said Haledjian, "the boy had some silly alibi about where he was when Miss Casey was slugged and robbed?"

"Right. He claims he got a note to be in the school boiler room at ten—fifteen minutes before Miss Casey was assaulted. He hung out for half an hour, but nobody turned up."

"I trust you made an arrest?" asked Haledjian.

Modified text used in the study including target items from the story "The case of the missing button". Target items are underlined. The expansion sentence is in italics.

Inspector Winters scrawled Matty's answer in his notebook and replied, "She isn't positive—she saw only your back. But this missing button proves you did it. Luckily, Miss Casey isn't badly hurt. Now, you don't want to bicker with me, do you? *Let's not fight.* Tell me, where's her bag?"

"Matty kept saying he didn't know a thing about the robbery. He said he salts away money from his job and wouldn't rob someone for money," the inspector told Dr. Halliday later.

"No doubt," said Halliday, "the boy had some silly story about where he was when Miss Casey was hit and robbed?"

Inspector Winters stifled a yawn. "Right. He claims he got a note to be in the school boiler room at ten o'clock—fifteen minutes before Miss Casey was robbed. Students almost never go to the boiler room because it is hot there. He frittered away half an hour, but nobody showed up."

"I guess you sussed out who did it. Are the papers drawn up? Let's nab the student fast." said Halliday.

Quantitative Measures: Overview. As referenced above in the breakdown of the nine levels of knowing a word in Nation (2001), knowledge of a word involves receptive and productive aspects covering form, meaning and use. Word knowledge ranges from the more shallow aspects of knowledge, such as word form recognition, to deeper aspects such as collocations or constraints on use. In order to test the effectiveness of reading as a means for incidental vocabulary acquisition to occur, it is necessary to test both receptive and productive knowledge. Accordingly, the tests chosen for the post- and delayed post-tests reflected both types of vocabulary knowledge within the three aspects outlined by Nation (form, meaning, and use). Pre-tests were used to collect data for target word selection and level checking of the participants. For all pre- and post-testing learners worked on one test at a time. In other words, when a learner completed a test, he/she handed it in to the teacher and received the next test. This allowed learners to complete the test in their own time. Any learners who finished all tests while others were still working were allowed to go on break.

Pre-Test 1: Self-Rated Protocol. The test was a self-rated protocol also employed in Horst (2005). In this protocol, a list of 200 words including phrasal verbs, verbs, nouns, adjectives and adverbs were presented to participants. Participants checked one of three choices for each word: YES (I know what this word means), NO (I do not know what this word means), or NS (I think I know what this word means but I am not sure). Nouns, adjectives and adverbs were included to help mask the intended target items. Most of the students marked target vocabulary items NO or NS; of the 16 targets, only three PVs and no one-word verbs were marked YES by ¾ of the students. Hence the selection of vocabulary described above appears to have succeeded in identifying words

learners did not already know for inclusion in the stories. This test assessed receptive knowledge only. A sample of 10 test items appears below (*pack in, bicker* and *swill down* are target words); the full test is available in Appendix B.

100. go back	___YES ___NOT SURE ___NO
101. holy	___YES ___NOT SURE ___NO
102. pack in	___YES ___NOT SURE ___NO
103. sit back	___YES ___NOT SURE ___NO
104. establish	___YES ___NOT SURE ___NO
105. tune	___YES ___NOT SURE ___NO
106. take up	___YES ___NOT SURE ___NO
107. bicker	___YES ___NOT SURE ___NO
108. justify	___YES ___NOT SURE ___NO
109. scared	___YES ___NOT SURE ___NO
110. swill down	___YES ___NOT SURE ___NO

Figure 1: Self-rated pre-test example

Pre-Test 2: Nation's Levels Test. Participants were tested on the receptive and productive 1,000 to 5,000 levels of Nation's (2001) Levels Test since, barring the target words, the stories contained words from among the 2,000 level. This test was used to ascertain whether all participants were at roughly the same level of vocabulary knowledge breadth. As expected, most intermediate students tested at roughly the same level with the advanced students showing greater vocabulary knowledge breadth. The mean score on the section of the test that samples knowledge at the 2,000 frequency level

was 26.69 (*SD* 2.94) for the receptive section and 11.56 (*SD* 3.41) for the productive section; a maximum score of 30 was possible for each of the receptive and productive tests for each frequency level. The mean score on the 3,000 frequency level was 21.5 of a possible 30 (*SD* 5.67) for the receptive test and 6.13 of a possible 30 (*SD* 5.16) for the receptive test. Thus, it can be assumed the stories were comprehensible to the readers in terms of their known-word content.

Post-test Receptive Test 1: Self-Rated Protocol. Basic form and meaning recognition was tested using the same self-rated protocol as the pre-test in order to judge if the items marked NO or NS on the pre-test had moved to the YES column. Question items were re-ordered in the post- and delayed post-test versions.

Post-test Receptive Test 2: Multiple Choice Definition Match. The multiple choice definition match test was modeled on the Nation's Levels test as described above in the pilot study section. Students were given 10 questions to answer. Each question contained six items consisting of one target phrasal verb, one target one-word verb, and four distracters. Three definitions were provided for matching the items to, with two of the definitions matching the two targets. This test measured ability to recognize word forms and their definitions. The entire test can be seen in Appendix C.

Post-test Productive Test: Sentence Writing. As in the pilot study, participants were prompted to write a sentence using a specific verb. A prompt was given with the verb to be used in brackets at the end, as in the pilot study example above. For phrasal verbs, the depth of the form-meaning connection made was tested by only providing the verb headword and not the collocating particle thus requiring students to provide the particle in their answer. The entire test can be seen in Appendix D.

Procedures

Instructional Treatment. The experiment took place during eight classes of 90 minutes each over a period of four weeks. The instructional treatment occurred during five of the eight research classes; only the four classes falling between days two to five presented targets in the stories. Table 1 outlines the experimental schedule.

Table 1

Experimental Schedule of Classes

Class Number	Week	Class Description
1	1	Pre-test and biographical data collection
2	1	Instructional treatment story 1 & 2
3	2	Instructional treatment story 3 & 4
4	2	Instructional treatment story 5 & 6
5	3	Instructional treatment story 7 & 8
6	3	Post-tests
7	4	Two decoy instructional stories
8	4	Delayed post-tests

All students underwent the same instructional treatments, testing, and initial biographical information gathering. Each instructional class was split into three segments of time with a story read in the first and last segment. During the story segments participants discussed two global warm-up questions, read along while a one-page mystery story was read aloud by the teacher, and then proceeded to try and solve the mystery. Learners were given 10 minutes to find a solution through discussion in pairs, which was recorded for later analysis. This was followed by a five-minute class discussion to reveal the mystery's solution which was also recorded. As mentioned above, the oral discussion component of the experiment was likely to increase exposure through oral use of the target items. Time on task and exposure to the written targets were carefully controlled. All students who guessed the correct solution were given a candy. Table 2 outlines the instructional class plan.

Table 2
Instructional Class Plan

Step	Time	Description
1	10 mins.	Warm up discussion on topic related to story (eg. jealousy, outdoor activities, etc.)
2	2 mins.	Paper copy of the story was distributed
3	5 mins.	Story was read aloud by the teacher
4	10 mins.	Pairs of students were recorded discussing the possible solution to the mystery
5	5 mins.	Short class discussion on the solution to the mystery, also recorded
6	10 mins.	Short unrelated communicative activity
7	40 mins.	Steps 1-8 were repeated for the second story

All activities were carefully designed to not include any targets in the questions or instructions. Although the teacher took care not to use the target items at any time, one target, *nab*, was inadvertently used once by the teacher when addressing the class.

The stories were read aloud to ensure that any learners who knew the sound of a word but not the written form would be able to make the form-meaning connection and hence increase understanding of the story. Reading aloud also ensured that the stories were read in their entirety and at the same rate by all learners (following a method used by Horst, Cobb & Meara, 1998).

Unrelated communicative tasks, such as surveys and unrelated communicative discussion activities, were carried out between the readings of the two stories. The activities were intended to help clear the target items from the students' minds (Hulstijn, 2003) and also gave a cognitive break so students would not be overwhelmed by reading and solving two stories in a row.

As in the pilot study, any questions about vocabulary items were not answered by the teacher but redirected to other students. Asking questions about unknown words meant that any questions about targets increased frequency of exposure and salience of those specific targets for all students within earshot. If any student asked about an item, the

item and student's name were recorded for use in answering the third research question about student questions increasing salience. This data was collected from the recordings made during discussions. Dictionaries were not allowed in order to help ensure any incidental learning occurred from the in-class reading treatment only and not other influences. Participants did not receive copies of the stories to study outside of the class.

Testing Schedule. Due to class scheduling, long breaks between the post- and delayed post-tests were not possible but it was hoped that the one week which elapsed was sufficient to claim the item had entered long-term memory (Hulstijn, 2003). The first treatment class occurred two days after the pre-test was administered. The post-tests occurred at spaced intervals. The immediate post-test occurred two days after the last treatment class. Although it appears less than ideal, this two day break mimics real-life learning where students may not be required to immediately recall and use an item they were incidentally exposed to. The delayed post-test occurred one week after the post-test with one intervening reading treatment class between; the intervening class texts did not contain any target words. Due to the week delay between testing, the same measures were used for both the immediate and delayed post-test although the order of questions was changed on each test.

Table 3
Order of Post-Tests for Immediate and Delayed Post-tests

Test Order	Test Description
1	Self-rated Protocol
2	Multiple Choice Definition Match
3	Sentence Writing

Recorded Data Collection. Recordings were made during pair and class discussions to solve the mystery. These recordings were reviewed later to identify how the target items

were negotiated, whether discussed for meaning, used naturally during discussion, or overheard, as well as to count the number of negotiations of each target item.

Analysis

Answering research question 1, which considered whether incidental acquisition would occur from eight repetitions of the targets in the written text, began with a raw count of the YES (I know this word) answers for each student on the self-rated pre-, post- and delayed post-tests for the 16 target PVs and one-word verbs. Evidence of a form+meaning connection being made was measured by changes in numbers of YES answers on the self-rated test. The means of the numbers of YES (I know this word) answers were calculated for each test across the experiment. To confirm the patterns suggested by the means, a repeated measures ANOVA was used. Post-hoc *t*-tests were used to identify whether the differences in the data for the two types of words were statistically significant.

The analysis of research question 2, which addressed the issue of whether the two types of verbs, PVs and one-word verbs, were learned to the same extent in the experiment, began with removing target items marked YES (I know this word) by 75% or more of students from the analysis. Only three target items were marked as being this well known; surprisingly, all three were PVs (*draw up, spell out, pack in*). These items were subsequently removed from the following calculations.

The next step involved two independent qualified ESL teachers who marked the sentence writing tests resulting in an average score for each student's sentence writing answers. The marking scheme is as follows: a full mark of 1 was awarded for grammatically and semantically correct use of the target word; half marks of 0.5 were awarded for correct semantic use but incorrect syntactic use, such as incorrect placement

of the phrasal verb particle; and a score of 0 was given for incorrect semantic and syntactic use. In order to enter a single score into the analysis, rater mismatches were resolved by taking the average of the score assigned by both raters. Interrater reliability was 80%. Examples of student writing from the post-test follow that show the various marks assigned.

Full mark of 1 awarded:
Phrasal verb: suss out
Question: What are police detectives good at? (suss)
They are good at sussing out who is the real criminal.

One-word verb: scrawl
Question: How would people describe your handwriting? (scrawl)
People generally said that I scrawl when I write.

Half mark of 0.5 awarded:
Phrasal verb: buck up
Question: Your friend calls and is upset, what do you do? (buck)
*I try to stay quite and buck up with her.

One-word verb: flaunt
Question: What would you do if you won two million dollars? (flaunt)
*I wouldn't flaunt about it.

Incorrect mark of 0 awarded:
Phrasal verb: swill down
Question: What do some people do with their favourite drink? (swill)
*They swill it out.

NOTE: The student correctly uses a particle although the wrong one.

One-word verb: pelt
Question: What month does it rain the hardest? (pelt)
*I don't know, I have to pelt out.

NOTE: Interestingly many students added the preposition *out* to their answers for this question.

The sentence writing scores were used to calculate accuracy percentages for PVs and single word verbs for each student. Accuracy percentages for each question were calculated by counting the number of sentence writing questions with a score of 0.75 or 1, representing the combined score by the raters, divided by the remaining target items of each type of verb (five PVs and eight one-word verbs, respectively). *T*-tests for matched samples were then used to test for differences in performance on the two types of verb.

The final research question dealing with the interaction of the incidental written exposures and oral negotiation was answered by doing raw counts followed by a standard correlation to test for relationships between the exposures and negotiations. Oral negotiation included any items the learners used or (over)heard another person use; overhearing was judged by the researcher as the target item use being audible on the recording of the discussion. All targets indicated as previously known on the self-rated pre-test were removed per student case. The first raw count was of items each of the 14 participants marked unknown (NO or NS) on the self-rated pre-test which were subsequently negotiated (used or heard by the learner). These items were then compared to the multiple choice and sentence writing answer data to find counts for items answered correctly and incorrectly on the multiple choice and sentence writing test respectively for both the post- and delayed post-tests. To look at these data more closely, a standard correlation was used where each negotiated item was correlated to the number of correct answers for the post and delayed post multiple choice and sentence writing tests, respectively. A second standard correlation was run correlating negotiated incorrect answers for both tests at the two time intervals.

Chapter 3: Results

Results of the experiment are presented in the following order. First, the question of whether the learning activities (consisting of eight written reading encounters with the target words and possible additional encounters during the mystery-solving) resulted in measurable amounts of new word learning is addressed. Then the second question, which considers whether phrasal verbs and one-word verbs were learned differentially, will be answered. Finally, the last section will report findings that address the third research question about the effects of oral interaction on acquisition.

Research Question 1. Answering the question about whether eight contextual written exposures and comprehension-focused follow-up activities resulted in acquisition of the target items involved tallying numbers of YES (I know this word) answers on the self-rated test over the pre-, post-, and delayed post-tests. Group means for the two types of word targets at all three measurement points are shown in Table 4. The figures show PVs were better known at the outset than the single words; means for both types of items rise in value across the testing times pointing to knowledge increasing over time. The data in Table 4 are shown graphically in Figure 2.

Table 4
Means of YES Answers on Self-rated Test Over Time (n=14); Maximum = 8

	PVs			One-word verbs		
	Pre-test	Post-test	Delayed post-test	Pre-test	Post-test	Delayed post-test
M	3.28	5.14	5.64	0.86	2.86	3.86
SD	1.12	1.88	1.50	0.86	2.14	1.82

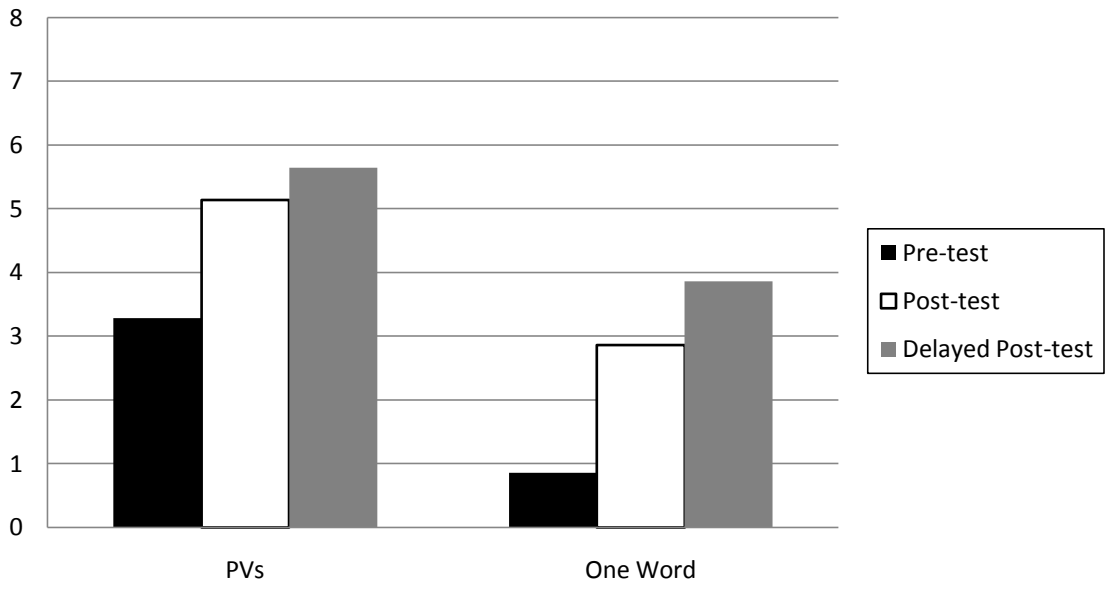


Figure 2: Means of YES Answers on Self-rated Test over Time

A repeated measures ANOVA indicated that there were significant differences in learner performance on the two types of tests ($df = 1, 27$, $F = 18.39$, $p < 0.001$) and over time ($df = 2, 56$, $F = 34.94$, $p < 0.0001$) but no significant interaction. That is, performing well on one test, such as the PVs, did not predict or assist performance on the other test and vice versa. Post-hoc HSD Tukey tests showed that pre-test and post-test means for the PVs differed significantly ($p < 0.05$). The difference between the pre-test and delayed post-test means on this measure was also statistically significant ($p < 0.05$). However, although Figure 1 shows an increase in mean numbers of PVs rated YES over the week between the post-test and delayed post-test, this difference did not prove to be significant. A similar pattern of results was found in the case of the single-word items indicating that knowledge of these items increased and that gains were maintained over time: the pre- and post-test difference in means was statistically significant as was the difference between pre- and delayed post-test means. But again the difference between the post-test

and the delayed post-test mean was not significant (even though Figure 2 suggests that the difference was substantial). In summary, knowledge of both types of lexical item appears to have improved as a result of the experimental learning activities and gains appear to be stable. Of the eight PVs assessed, the mean number rated YES before the learning activities amounted to about three; this increased to almost six by the end of the experiment for a mean gain of about three PVs. The single words means also increased by three items, with about one rated YES at the outset and nearly four by the end. Thus, the hypothesis that eight contextual written exposures plus any exposures during the oral interactions would lead to gains in numbers of new form+meaning connections was confirmed.

So far, the findings suggest that amounts of growth were similar for the two types of lexical items, but the comparison is somewhat limited by the fact there were different starting points, with PVs appearing to be better known at the outset of the experiment than the one-word verbs. It is also possible that learners did not self-rate their knowledge of the targets accurately. Outcomes based on measures that allow for more reliable comparisons of knowledge gains are the topic of the next section.

Research Question 2. The second research question investigated whether eight contextual written exposures and comprehension-focused follow-up activities would result in differential depth of knowledge for PVs and one-word verbs. Comparisons based on two measures, one that tested productive ability and another that assessed recognition knowledge, are discussed in this section. In order to be able to make more valid comparisons, three PVs (*spell out*, *draw up* and *pack in*) that were previously known, having been marked as YES by ¾ of the participants, were left out of the analyses.

Results based on the five remaining PVs and eight one-word verbs (all of which were unknown to the participants at the outset of the study) are reported in percentages to allow for easy comparisons across the two kinds of lexis. Some acquisition of both types of items was expected to occur given previous studies of incidental vocabulary acquisition through reading; of particular interest here was any evidence that the more complex two-word targets (PVs) were more difficult to acquire. First, performance on the sentence writing test is reported; this productive measure was challenging to complete and therefore more likely to be sensitive to difficulties presented by PVs than the receptive measure. Learner productions such as the following from the post-test illustrate these challenges:

Target PV: *suss out*
Question: What are police detectives good at? (suss)
*They are good in sussing criminals.

Target PV: *fritter away*
Question: Do you always like to be busy? (fritter)
*No, I like to fritter away on the street.

However, there is also ample evidence of ability to use the PVs correctly as the following post-test productions show.

Target PV: *salt away*
Question: Where is the best place to keep money? (salt)
I salt away some money in bank for rainy season.

Target PV: *buck up*
Question: Your friend calls and is upset, what do you do? (buck)
I buck him up to make him feel comfortable.

By the end of the experimental treatment, the findings shown in Table 5 indicate substantial gains in knowledge for both kinds of verbs on the sentence writing test, giving

evidence that deep learning occurred. Learners were able to use about 40% of the previously unknown PVs and single words in accurate sentences. Interestingly, this knowledge appears to have increased slightly by the time the delayed measure was administered again seven days later. The lower accuracy scores for PVs on both the immediate and delayed post-tests suggest that they were indeed more difficult to use in sentences than the one-word verbs. However *t*-tests for matched samples revealed that differences in the means for the types of lexis were not statistically significant ($p > 0.05$).

Table 5
Sentence Use of Previously Unknown Words by Word Type in % (n=14)

	Post-test		Delayed post-test	
	Phrasal verbs	Single verbs	Phrasal verbs	Single verbs
M	38.57	43.30	45.00	53.35
SD	26.42	18.50	24.81	20.89

As expected, mean performance on the multiple-choice measure indicates greater gains at the form+meaning level than the more challenging productive measure. The percentages suggest that after the experimental treatment, learners were generally able to use almost half the words they had previously rated as unknown in correct sentences (Table 5) while a form+meaning match was generally possible with half or more words (Table 6). Here the figures suggest that meanings of the one-word verbs were easier to recognize. However, *t*-tests for matched samples indicated that the differences in the means for the two types of lexis were not statistically significant ($p > 0.05$).

Table 6
Multiple Choice Answers of Previously Unknown Words by Word Type in % (n=14)

	Post-test		Delayed post-test	
	Phrasal verbs	Single verbs	Phrasal verbs	Single verbs
M	52.86	50.00	70.00	58.93
SD	34.74	27.74	32.11	23.73

In summary the experimentation showed that eight contextual written exposures and comprehension-focused follow up activities resulted in an almost equal amount of learning for both PVs and single word verbs as indicated by correct answers on the multiple choice test and the measure of ability to use the target words in sentences. Thus Hypothesis 2 that single word verbs would be learned more deeply than PVs was not confirmed.

Research question 3: Research question 3 investigated whether there is evidence that targets learners ask about are acquired more deeply than words that are not given this kind of attention. In answering this question, we considered any items the learners negotiated by using the item themselves in any way or by hearing someone nearby use the item as judged by the word being picked up by the microphone during recording and audible to the researcher upon later review of the tape. Phrasal verbs and single words were taken together in this analysis.

Both PVs and single word verbs were negotiated an average of 11 times each over the course of the four treatment classes. Each of the four treatment classes had a total of 30 minutes for discussion (two blocks of 15 minutes) in a 90-minute class; this means a total of 120 minutes were available for negotiation. Hence, with an average of 11 negotiations each, a target was negotiated in some way every 12 minutes. Table 7 shows total negotiations were almost equal for each type of verb with PVs negotiated 89 times and one-word verbs 88 times. A more pronounced difference is seen when only items that had been marked as unknown by the negotiators are counted (shown in the second row of Table 7). As the figures show, previously unknown PVs were negotiated 55 times and one-word verbs 78 times.

Table 7
Negotiations by Word Type in Raw Counts (n=14)

	Phrasal Verbs	One-word Verbs
Total Negotiations	89	88
Negotiations of Unknown Items	55	78

Table 8
Phrasal Verb Negotiations in Raw Counts (n=14)

	Draw up	Spell out	Suss out	Salt away	Buck up	Pack in	Swill down	Fritter away
Total Negotiations	11	9	14	13	12	12	4	13
Negotiations of Unknown Items	2	4	14	11	7	1	4	12

Table 9
One-word Verb Negotiations in Raw Counts (n=14)

	Pelt	Nab	Bicker	Flaunt	Fleck	Flinch	Stifle	Scrawl
Total Negotiations	14	14	9	10	14	10	3	14
Negotiations of Unknown Items	14	11	7	8	13	9	3	13

It is evident from Table 8 that several PVs that were not initially well known in the group were negotiated multiple times; examples are *suss out, salt away*, and *fritter away*. As Table 9 shows, unknown one-word verbs were negotiated much more frequently, with *pelt, nab, fleck* and *scrawl* receiving a great deal of attention. The following excerpt shows a successful negotiation of the phrasal verb *draw up* by three students where it is actively used five times.

> Student 1: But I think have the papers drawn up. Drawn up is mean, uh, write on the papers?
> Student 2: Draw drawn up? Maybe...
> Student 1: Write?
> Student 2: Maybe... drawn up
> Student 1: So, it's drawn up. Write something on the papers.
> Student 3: Yes, yes, yes, to prepare something.
> Student 1: If nothing is drawn up on the paper how can he tell that he got a note?

Students 1 and 2 had marked *draw up* as known on the self-rated pre-test yet they appear to not know the meaning which was provided by Student 3 who had marked it as

unknown on the self-rated pre-test. Only Student 3 was not able to make the form+meaning connection for the post-test multiple choice measure and all three scored full marks for *draw up* on the delayed multiple choice measure.

Negotiations did not always lead to clear inferences of word meanings as the following excerpt shows in a negotiation of the phrasal verb *fritter away* by the two strongest students. Student 4 had marked *fritter away* as unknown on the self-rated pre-test but Student 5 had marked it known. Hence, Student 5 had surface level recognition of the word and possibly some level of form+meaning connection already made which is evident in the transcript of the negotiation.

> Student 4: See, he said "Afterwards the whole group drove out here in several car, okay, to eat hot dog and fritter". What is fritter?
> Student 5: Fritter away?
> Student 4: Fritter and eat hot dogs and fritter away time. No, no, no, I think is...
> Student 5: Yeah, the group go to eat something and wander the the...
> Student 4: Yes... Fritter... Fritter away time...
> Student 5: Fritter away may be some, spend some time at Benny's restaurant.
> Student 4: Uh, cause eaten... uh, I think fritter is not...I think fritter is another thing...fritter away

Note that Student 4 initially does not recognize *fritter* as belonging to the particle *away* and by the end of the exchange is unable to accept the possible definition provided by Student 5. Student 4 repeats the verb headword and the entire PV several times but finally finishes by rejecting the use of the collocating particle although he repeats the correct PV once more before moving on. On the post-test multiple choice measure, Student 4 did not correctly match *fritter away* to its definition but was able to do so on the delayed post-test; similar results were found for Student 4 on the sentence writing measure.

To determine the possible impact of amounts of negotiation on learning outcomes, acquisition and negotiation scores were tabulated for each of the 16 target words (PVs and single words taken together). Acquisition was measured in terms of the number of correct and incorrect answers for negotiated and non-negotiated targets on the multiple choice and sentence writing post- and delayed post-test measures. Thus, for example, the PV *buck up* was negotiated by 12 students. Then any students who marked this PV as known on the pre-test were removed from the analysis leaving seven cases of negotiation by learners who did not already know *buck up*. Next the performance of each of these students for the item *buck up* on each measure was tabulated and entered into the schemes shown in Tables 10 and 11. In the case of *buck up*, two of the seven students identified the meaning correctly on the multiple choice measure. This figure was then entered into the count for "targets negotiated and correctly identified" (see the first row of Table 10). The sentence writing post-test figures in Table 10 show many more correct answers for negotiated (43) than non-negotiated (9) targets. However, on the challenging sentence writing test, the incorrect answers for negotiated targets (90) far outweigh the correct answers for negotiated items (43). Thus negotiating appears to be of some use but it is hardly a guarantee of arriving at accurate knowledge at the level of ability to use new words in sentences. Table 11 shows performance on the multiple choice recognition measure. Generally, numbers of correct answers are higher on this less demanding recognition measure. As was found for the sentence results shown in Table 10, there were many more correct answers for negotiated (69) than for non-negotiated (16) items. However, in contrast to performance on the sentence task, negotiation appears to have led to slightly more correct (69) than incorrect (64) responses. In other words, these findings

suggest that negotiation is more beneficial for knowledge at the level of meaning recognition than the sentence use level (though it is still clearly no guarantee that the learning will be accurate).

Table 10
Negotiated and Non-Negotiated Items and Sentence Writing Test Accuracy (n=14) in Raw Counts

Description	Post-test	Delayed post-test
targets negotiated and used correctly	43	58
targets negotiated and used incorrectly	90	75
targets not negotiated and used correctly	9	13
targets not negotiated and used incorrectly	24	20

Table 11
Negotiated and Non-Negotiated Items and Multiple Choice Test Accuracy (n=14) in Raw Counts

Description	Post-test	Delayed post-test
targets negotiated and correctly identified	69	82
targets negotiated and incorrectly identified	64	51
targets not negotiated and correctly identified	16	22
targets not negotiated and incorrectly identified	17	11

To further investigate the impact of negotiation of the targets on learning outcomes, correlations between amounts of negotiation and the accuracy scores were tested. Each of the 16 targets was assigned a negotiation score based on the number of times it had been discussed by a learner (or heard in a negotiation with a partner or nearby pair). The Pearson product-moment correlation between these figures and the numbers of correct answers learners provided for each of the 16 targets on the multiple choice post-test was found to be strong and significant ($r = 0.84$, $df = 14$, $p < 0.0001$). However, this promising result is compromised by the finding that the correlation between negotiation score and numbers of incorrect answers on the multiple-choice measures proved to be fairly strong ($r = 0.67$, $df = 14$, $p = 0.0022$). The same pattern of

results was found when negotiation scores were correlated with the delayed post-test multiple-choice scores and with the post-test and delayed post-test sentence scores. In other words, it appears that large amounts of negotiation are strongly and consistently associated with large numbers of both correct and incorrect answers on measures of target word knowledge. Perhaps negotiation inspired a sense of familiarity and gave the students false confidence in their knowledge. Hypothesis 3 that items learners ask about will be acquired more deeply was partially confirmed.

Conclusion. From the findings reported above, it appears that incidental exposure to phrasal verbs through eight unique written exposures and oral interaction does lead to acquisition. Gains were seen the most in the receptive realms of surface-level self-rated recognition and form+meaning recognition, with weaker levels of acquisition shown in the productive aspect of using the target to write a sentence. One-word verbs were acquired slightly more than phrasal verbs, although this difference was not statistically significant. This shows that incidental exposure through reading is effective for both kinds of verbs. Surprisingly, oral interaction proved useful to predict both correct and incorrect learning, which was hypothesized to stem from false confidence gained by using or hearing the word aloud.

Chapter 4: Discussion and Limitations

Discussion

This study has contributed to the larger body of research on incidental vocabulary acquisition and second language vocabulary acquisition through reading by looking at the incidental acquisition of one particular type of lexis, the phrasal verb, contrasted to one-word verbs. At the time of writing, no other vocabulary acquisition study was found which contrasted incidental formulaic sequence acquisition with the acquisition of one-word items through reading (Bishop (2004) used saliency when comparing items, which does not create an incidental learning environment). Previous research has shown incidental exposure is useful for formulaic sequence acquisition (Schmitt et al., 2004). However, except for Bishop (2004), since the previous research did not make comparisons between single words and formulaic sequences, the findings reported here present new contributions to the area of second language vocabulary acquisition through reading.

As hypothesis 1 proposed, incidental exposure to eight unique contextual occurrences in writing and oral follow-up activities did lead to gains in knowledge of both phrasal verbs and one-word verbs. This supports previous research findings suggesting that eight or more incidental reading exposures (Horst, Cobb, & Meara, 1998) are likely to lead to acquisition of new form+meaning associations. The learners' self-ratings indicated mean knowledge of PVs in the group increased from about three at the outset to about five at the post-test. While many studies indicate a decline in knowledge over time, this study found the gains were maintained and even approached six (almost double the original mean) on the delayed post-test, showing knowledge may have even been increasing. A similar pattern was found for the means of the one-word verbs. At the

outset, the group means for known items was less than one; this figure doubled at the post-test to almost three and the mean rose to almost four on the delayed post-test. The gains in learning from the post-test to the delayed post-test, although not statistically significant, may be indicative of continued growth; however, these gains may not have resulted from the reading treatments and discussion but instead have come from discussion of the tests and materials during the students' regular daytime classes.

While the self-rated tests showed changes in surface recognition, the measures that required learners to demonstrate their knowledge (the sentence writing and multiple choice measures) indicated learning occurred on deeper levels and helped confirm growth was indeed substantial. The uptake levels from this study are much higher in contrast to previous research of word learning through reading. As noted in the literature review, uptake rates of one tenth of the target words are typical in studies of incidental vocabulary acquisition (Waring & Nation, 2004). In this thesis study, learners were able to accurately use almost 40 percent of previously unknown words in sentences and correctly recognize the meanings of more than half on the multiple choice measure. Similar patterns were found for the one-word verbs. All three measures showed the gains were maintained over time with a surprising upward trend on the delayed post-test, which is opposite to the expected decline in knowledge usually observed. Given the findings of earlier studies, the uptake rate of 40 to 50 percent of items is impressive. This suggests that the combination of frequent encounters with new words through reading and opportunities to negotiate the reading material provide a highly beneficial context for vocabulary learning.

Hypothesis 2 predicted that the reading treatment would result in unknown one-word verbs being acquired more deeply than unknown phrasal verbs. It is interesting to note that significant gains occurred for both types of verbs on the sentence writing tests and that no significant differences between the two types of verbs were found. It was previously assumed that knowledge of phrasal verbs would be difficult to acquire but these findings tell a different story. The sentence writing task supplied only the verb headword of the phrasal verb and required the student to provide, and correctly use, the particle in the sentence they wrote. For example, a phrasal verb sentence prompt was "Do you always like to be busy? (fritter)". The student was required to supply the entire phrasal verb *fritter away* in a grammatically correct sentence for full marks. In this research, the sentence writing test indicated that eight incidental exposures through written text plus oral negotiation may be useful for acquiring single word items as well as potentially difficult multi-word formulaic sequences, such as the verb+particle structure of phrasal verbs.

The investigation of research question 2 also showed gains over time for both types of verbs on each measure. Learners were better able to make form+meaning connections giving rise to more correct answers on the multiple choice tests (as shown in Table 6) than the sentence writing tests (Table 5). These positive gains in incremental knowledge were expected. As often mentioned in the literature, these results confirm the pattern of incremental knowledge gains moving from receptive to productive knowledge (Nation, 2001). Incremental knowledge is measured by Paribakht and Wesche's (1997) Vocabulary Knowledge Scale (VKS) which measured for incremental knowledge by requiring learners to provide a synonym or translation for an item and if possible use the

word in a sentence marked for semantic and grammatical accuracy. Tests used in the thesis functioned in a similar way. The surface level recognition test, the self-rated test, occurred first. The multiple choice test, conducted as the second of the three tests, asked learners to choose the correct meaning option for a target word from among six distracters. The more difficult sentence writing test, completed last, required productive use of grammatical and syntactical knowledge of unknown targets and was marked based upon whether the learner has used the item completely correctly or simply demonstrated semantic knowledge. For example, the phrasal verb target *swill down* was marked as known by only one student on the self-rated pre-test. After the treatment, most students correctly matched *swill down* to its definition on the multiple choice measure, illustrating incremental knowledge gains. However, the sentence writing measure shows that the deeper aspects were more difficult for students to acquire. Following are example sentences and the grades received:

> Sentence prompt: What do some people do with their favourite drink? (swill)
> Full marks (minor grammatical errors were ignored as per Paribakht and Wesche (1993, 1997):
> They'll swill down their drink.
> They will swill down quickly.
>
> Half marks:
> Usually people drink their favourite drink but same time they can even swill down by it.
>
> Incorrect:
> My son swilled down pizza because he liked it.
> Note: This student has hypothesized, although incorrectly, that *swill down* can be used with food as well as beverages.

As mentioned above, the increase on the delayed post-test scores may have been due to continued student discussions of the targets outside the research setting.

The effect of possibly increasing the saliency of unknown items through negotiation coupled with incidental written exposures of unknown vocabulary was investigated in research question 3. The expectation of hypothesis 3 that negotiated items would show higher gains than non-negotiated items was borne out by the greater numbers of correct answers for negotiated items on the post and delayed test measures. In other words, either using an unknown word or overhearing another learner use an unknown word increased the likelihood the learner would acquire some level of knowledge of the unknown target.

It is important to point out that in order to maximize reading comprehension each text was carefully controlled to ensure that unknown items counted as only five percent or less of the total word count per text (Laufer, 1992; Laufer & Ravenhorst-Kalovski, 2010). The small number of unknown words would have made them particularly salient for negotiation of meaning or for silent contemplation. Yet, targets were not frequently negotiated with only one negotiation roughly every 12 minutes. Since the material did not require use of the targets in order to solve the mystery and no comprehension questions had been asked, the amount of negotiations were not as great as they might have been in more typical reading-related interactions. Since the number of previously unknown single word verbs were negotiated more (78 times) than PVs (55 times), it seems the potentially more difficult phrasal verb may have resulted in some oral use avoidance (Dagut & Laufer, 1985; Hulstijn & Marchena, 1989; Laufer & Eliasson, 1993) while learners were more comfortable testing hypotheses with new one-word verbs. When known targets were included in the analyses, the two target types were used almost equally (PVs 89 times, single word verbs 88 times).

Surprisingly, testing of the correlation between amounts of negotiation and acquisition also revealed classroom interaction can predict both accurate and inaccurate knowledge. More negotiation did not always lead to more accurate acquisition and a lack of instructed feedback forced students to rely on their own accumulated knowledge, hypotheses, or the knowledge offered by their peers. During a think-aloud protocol while writing a short text, Swain and Lapkin (1995) also found that young immersion students sometimes generated "incorrect hypotheses and inappropriate generalizations, suggesting that relevant feedback could play a crucial role in advancing their second language learning" (p. 384). Incorrect answers were as plentiful as correct answers on the sentence writing measure although unfortunately, the tests were not sensitive enough to measure why this was so. It is possible to say that both correct and incorrect learning can be predicted by negotiation; perhaps the incorrect knowledge may result from negotiation leading to false confidence. This possibility is discussed in more detail in the next chapter.

Negotiation does show usefulness by increasing the number of incidental exposures learners get to the vocabulary. As outlined in the literature review, negotiation increases the minimum number of eight encounters research has suggested are necessary for learning. Increased exposures can lead to deeper learning. In this study, negotiation of unknown items did lead to gains in knowledge as evidenced by correct answers even though lack of learning evidenced by the incorrect answers for negotiated items also occurred.

Finally, it is interesting to note that learners' professions combined with English proficiency may have aided in ability to solve the mystery story. Two of the low-

advanced learners who had previously worked as lawyers were particularly skilful at solving the mysteries. These two learners were good sports by allowing other students the opportunity to put forward their guesses of the solution before offering their own.

The next section will deal with limitations that were identified in this study.

Limitations

Three main areas are addressed in the discussion of limitations identified in this study: the context of the study, testing formats used, and the design of the materials.

Turning first to the research context, the students were drawn from a collection of daytime LINC classes occurring at the same centre. It was not possible to use a truly random sample of students and the students who answered the call for participants were accepted as an intact class. Although it is recognized as a limitation, it was necessary to accept the group in its entirety in order to have the largest possible sample size of participants.

The learner levels had been previously assessed by the LINC centre which placed the students between low-intermediate and low-advanced. Only two of the 16 students were in the low-advanced group. It was also observed by the researcher, an experienced ESL teacher, that two low-intermediate students were much weaker than others and seemed to border on high-beginner. The range in abilities was confirmed by performance on the Vocabulary Levels Test which showed a range of scores on the 2,000 level words from 18 to 30 (out of 30 marks) on the receptive measure and seven to 17 (out of 30 marks) on the productive measure. This disparity in level was unavoidable although the rest of the 14 participants appeared to be at the level reported by the LINC Centre. As often happens in classroom research, attrition of participants occurred. Originally 17

participants came for the class but one did not return for unascertainable reasons. Two of the remaining 16 participants had to be removed from the statistical analyses as they missed either the post- or delayed post-test due to unavoidable commitments which conflicted with the class time. It is recognized that 14 participants is a small sample size and accordingly, it is important results not be extrapolated onto the larger body of ESL learners until further research duplicating this study is carried out to confirm the statistical findings are consistent with larger numbers of participants.

No comparison group was included in this study, which is another recognized limitation. A comparison group of students who had not received exposure to the targets was not included since this type of comparison would not have shown the effectiveness of the treatment; the literature previously reviewed indicates incidental learning does occur and this study was not measuring if incidental learning occurs or not, but rather how to better use incidental learning in the classroom. A control group of students who did not receive treatment would only measure their existing knowledge from general exposure plus any previous non-measured instruction, which is not useful in showing gains when compared to a group that did receive treatment which was measured, as in this study. However, a comparison group would have shown if any test effects occurred between the post- and delayed post-tests. That is, it would have been possible to identify whether having the two post-tests only a week apart led to any learning that can be ascribed to the tests themselves, rather than to the experimental treatment. Any gains showed by the control group would indicate that gains made by the experimental group are explained at least in part by seeing the words (and their definitions) repeatedly on the two post-tests. A second comparison treatment group was indeed originally planned but

the idea was discarded. The second treatment group would have received the same reading treatment but instead of orally discussing the mystery solution would have been required to write three possible solutions to the mystery story. It was decided to drop this comparison group since students may have found the task too difficult and potentially boring, possibly resulting in a high attrition rate. As it was, the participants of this study who discussed the solution with a partner still found solving the mystery challenging. In addition, if the students were drawn from the same small Centre, the groups may have discussed what they were doing in the research class and the silent writing group may have been discouraged by the lack of discussion time, again possibly leading to attrition.

A second issue related to the context of the study is the possibility of learning that occurred outside the reading class. One of the students' daytime teachers often expressed to the researcher how much the students were enjoying the experimental class. This indicates the students were discussing the class, and potentially the material, during the regular daytime class, which would have resulted in uncontrolled and uncounted incidental exposures to the target words. The delayed post-tests showed a continued increase in correct answers which may have been due to discussions during the one-week lapse between the post-test and delayed post-test. While it was impossible to control such exposure, if the discussions occurring outside the experimental class focused on the test items (even though randomly ordered and with an effort to disguise them across the tests), the targets would have become more salient; if these discussions did indeed occur, it does lend weight to the benefits of oral interaction on salient items. If such discussions made the items more salient resulting in an increase in delayed post-test scores, this echoes Bishop (2004) where salient hypertexted formulaic sequences were looked up

more than non-salient items resulting in higher reading comprehension scores. Unfortunately it is impossible to query students as to whether discussion happened in the intervening week between the post-test measures in this study since students have moved on from the LINC Centre.

The next area of limitations deals with testing issues. It was impossible to judge whether items learners marked YES on the self-rated pre-test were actually known or were just mistaken for another phrasal verb they knew. For example, some students may have mistaken *pack in* for *pack up* since the latter was often used during the post-test sentence writing. Three phrasal verbs, *draw up, spell out*, and *pack in*, were marked YES by ¾ of the learners on the pre-test self-rated test and subsequently removed from the data used to answer Research Question 2, but it is possible that students did not know these items as well as they assumed. The productive post-measures that required demonstration of knowledge by recognizing correct definitions or producing accurate sentences (or both) were important for determining whether the surface recognition measured by the self-rated test was real or not. The pilot study had administered the productive tests during pre-testing but this was not duplicated in the thesis study as pilot study participants found the pre-test productive measures very taxing. Using the VKS where participants provide a definition and sentence if known would be a satisfactory way to measure knowledge before treatment and possibly avoid the issue of incorrect assumptions of knowledge by learners on the self-rated pre-test.

The multiple choice test showed the learners had knowledge of the correct form+meaning connections for words that they were not able to use well on the sentence writing test requiring deeper knowledge. Thus this less demanding recognition test was

useful in delineating stages of growth along the recognition-production continuum. However, guesswork may have also contributed to the higher scores on this measure. So, although the sentence writing scores were lower, they may be more accurate representations of knowledge gains by showing not only meaning comprehension but a grasp of the deeper syntactical and grammatical aspects of newly acquired lexical items, including the formulaic structure of the phrasal verbs.

The format of the sentence writing test presented two testing issues. For this test, depth of knowledge was judged by the correctness of a sentence learners wrote to answer a prompting question with a given verb, for example, for the PV *fritter away* students were given *Do you always like to be busy? (fritter)*. The first issue is that word knowledge between the targets was not tested equally. Learners were required to provide the associated PV particle as well as use it in a grammatically and semantically correct sentence, which wasn't required for the one-word target. The only way to avoid this difference would be to provide the whole PV. This is an option but data relating to how the PV is acquired, i.e. in stages or holistically, would be lost. The second issue relates to the information provided by the questions. For students with partially developed knowledge, providing the correct target beside the question may have helped consolidate their knowledge further since the question prompts unavoidably pointed to the definition of the target being tested. Other models were considered, such as grouping the targets at the top of the page with the question prompts below and instructions to use one word from the top of the page per question. Another option was providing a list of the targets without question prompts but this was decided against as students may have found it too difficult leading to increased testing time and frustration. The format used was decided

upon since a student who does not know the meaning of the word will still not be able to use it correctly in a sentence, as seen in the results such as the student answer for *salt away*, The best place to keep money is salting. In hindsight, the model with the words grouped at the top of the page appears to be better as it does not so easily link the target with the question prompt.

It is not possible to predict from this data how a learner's incorrect knowledge would change given more exposures, whether incidental or intentional and whether in written or oral form.

Another limitation of the measures was the small number of questions on the multiple choice and sentence writing tests with 10 and 16 questions respectively. The self-rated test had 200 test items including distracters; however, the ease of this test countered the large number of questions as students simply had to check if they knew, thought they knew, or didn't know a word. As discovered during the pilot study, the multiple choice item+definition match and sentence writing tasks were challenging for students. The pilot study showed including distracter questions would have helped mask the target questions better although this would have probably made the tests too long to complete within the space of a class. As it was, the three tests took most students about an hour to complete.

Furthermore, the assumption that a week delay between the post- and delayed post-test would be enough for students to forget which questions were asked on the post-test was proved questionable. Discussion with students at the end of the delayed post-test measure revealed that a few students had indeed discerned the target vocabulary from the re-ordered but repeated questions used again on the sentence writing delayed post-test.

This may have accounted somewhat for the maintenance and improvement of test scores on the delayed post-test. In order to avoid this limitation, future studies need to ensure that different question wordings are used for each target item on each test type over different test times.

The sentence writing tests were graded by two independent and experienced ESL teachers. Interrater reliability was 80%. More time for consultation between raters might have improved this reliability rating but was not available.

It is also recognized that testing a small number of words has associated limitations. The gains reported above appear very impressive, namely students were able to use about 40 percent of unknown items correctly in sentences and correctly match about 50 percent of items to a definition. However, this directly translates into a gain of only a few words which may not appear as impressive considering the thousands of words necessary for a proficient language user to learn. A gain of 50 percent meant a gain of only four words since there were a total of eight of each type of lexis. Using only eight PVs and eight one-word verbs was necessary due to the short length of the reading texts used in the treatment. The one-page mystery stories were chosen since a mystery gave learners a holistic goal for reading, beyond lexical comprehension, and previous unrelated teaching experience had shown longer mystery stories would have resulted in completing only one story per class instead of the two in the study reported here. Another limitation was the necessity of having the unknown target words represent only five percent or less of the 367 average total words used per story in order to ensure the texts were easily comprehendible (Laufer, 1992; Laufer & Ravenhorst-Kalovski, 2010). Having unknown items make up only two percent of the total words would have been

even more desirable (Nation, 2006; Schmitt et al., 2007; Laufer & Ravenhorst-Kalovski, 2010) but this would have meant an even smaller number of target items or longer texts, both of which were not feasible for reasons mentioned above.

The last overall set of limitations to discuss relates to the various alterations of the original material used to create the treatment texts. First, in order to add the target vocabulary to each of the eight treatment stories it was necessary to build in sentences containing the targets. However, these changes lessened the true authenticity of reading a natural English text. The alterations were unavoidable since most targets did not naturally occur in the texts. It is interesting to note that the three PVs, *draw up, spell out,* and *pack in*, which were identified as known on the self-rated pre-test by ¾ of the learners, were the only three targets which originally occurred in the texts and were used as targets.

Secondly, the recycling of each of the targets eight times through the texts also created an unnatural reading situation. Natural English texts do not tend to use the same set of vocabulary repeatedly over a number of texts or through the same text, unless there is a theme requiring this such as a series of related newspaper stories or the chapters in a novel. This recycling may have made the texts sound somewhat contrived.

Finally, since the intended readers were learners of English it was decided to grade the language in the texts in order to make them as accessible as possible. Grading took place on the sentence and word level. Sentence level changes consisted of building in story telling techniques such as signaling setting changes and adding speaker identification for the ease of ESL readers. Word level changes consisted of simplifications of words that were not among the 2,000 most frequent words in English as identified by the lexical profiling of each story with the BNC lists available on

www.lextutor.ca. Most words over the 2,000 most frequent were changed but in cases where words that were outside the 2,000 most frequent could not be easily substituted, appositive statements were required. Although every attempt was made to keep such changes to a minimum, the treatment stories cannot be called truly 'authentic' and the label 'graded text' would be more appropriate. However, students did not remark about any text idiosyncrasies to the teacher or to each other as audible on the recordings. The students told the teacher they enjoyed reading the stories, although the mysteries were challenging to solve. Perhaps the overall learning goal of solving the mystery successfully diverted some attention away from the repetitive vocabulary.

Having discussed the findings in terms of their strengths and weaknesses, I now turn to the discussion of the implications for research and language pedagogy.

Chapter 5: Implications for Research and Teaching

Implications for Research

There are several areas where further research could address either limitations of this study or use the findings of this study as a jumping off point for different research.

First of all, running this study again using not only more participants to validate the findings but also with comparison groups is needed. Comparing two groups, one who discussed the solutions as in this study and one who silently write possible solutions, would give fascinating insight into the effect of negotiation on vocabulary acquisition. Unfortunately, in either case it is impossible to determine how many times learners are incidentally exposed to the written targets during the time allotted for solving the mystery unless eye measurement equipment is employed.

The possibility of using eye measurement equipment to count how many times learners look at certain words in the text would be an interesting study in and of itself. If a group of learners doing the writing treatment, with no negotiation to solve the mystery, were measured in such a way, it may shed light on how many times unknown phrasal verbs are read and re-read in the text in comparison with unknown one-word items. The eye movement counts and post-treatment measures would illustrate how effective the mystery story task is for vocabulary acquisition of formulaic sequences as a representation of presenting incidental vocabulary learning opportunities through reading activities without intentional learning aims.

The eight incidental exposures measured in the investigation of research question 1 were chosen as a baseline figure from previous research (Horst, Cobb, & Meara, 1998) as a minimum number of incidental exposures for acquisition to occur. The literature

review referenced the ongoing discussion as to what is the optimal number of incidental exposures for acquisition to occur (Waring & Nation, 2004); by contrast the minimum threshold of known words in a text is generally agreed upon at 98% (Nation, 2006; Schmitt et al., March, 2008). The lengths of the texts used in this study necessitated having closer to five percent of unknown words in the text (Laufer, 1992; Laufer & Ravenhorst-Kalovski, 2010). An improvement on this study would be to use the same eight incidental exposures with longer texts in order to better mask the target vocabulary. Using longer texts would require more reading and discussion time at each stage of the experiment. Since the experimental treatment occurred across four sequential classes, it is possible the recycling of the same vocabulary in the short texts helped learners to, consciously or subconsciously, notice the targets. Longer texts would help circumvent this possibility.

Since the structure of this experiment showed gains in vocabulary acquisition at a rate of eight unique incidental textual exposures plus negotiation, it would be interesting to restructure the experiment to increase the exposure rate to possibly 18 following findings suggested by Waring and Takaki (2003). This could be done by using more individual texts of a longer length over a longer experimental period. One semester or more with an intact reading class would be an ideal timeframe for such an undertaking.

Although the finding that PVs and one-word verbs were not acquired differentially as expected but instead exhibited relatively similar gains (research question 2) appears to be good news; it is probably premature to conclude that all formulaic sequences are acquired as easily as one-word items. This is an interesting area for further study. Perhaps the question should be investigated in an experiment that mixes the

grammatical categories of formulaic sequences and one-word items being compared. Since phrasal verbs and one-word verbs belong to the word class, this may explain the almost equal levels of acquisition; further research is needed comparing different lexical items, such as a one-word verb and a relative pronoun formulaic sequence like *in the meantime*.

Learner responses on the sentence task point to an interesting avenue for future research. On answers scored 0, an incorrect particle was often supplied by weaker and stronger students alike showing the possible beginning of a learning process that starts with associating the correct meaning for the main verb. For example, with the phrasal verb *swill down* a student wrote "They swill it out." Perhaps the ability to supply the particle develops later. This indicates a possible area where future research can investigate whether phrasal verbs are acquired in stages or as a complete unit. If they are acquired in stages it would be interesting to examine the speed at which learners move through the stages or if factors like the particle itself, *out* versus *in* for example, play a part or, in addition, the possible impact of the verb headword itself as high- or low-frequency, for example the low-frequency *swill* (BNC 14,000 word frequency) versus the more common *pack* (BNC 1,000 word frequency).

With the one-word verbs students sometimes provided an extraneous preposition abutting the verb in their sentence. Many students unexpectedly provided the particle *out* when using the one-word verb *pelt* as if mistaking it for a phrasal verb. It is unclear why this occurred and it is an interesting area to look into how the number of phrasal verbs in a text could influence learners' acquisition of the one-word verb forms.

As seen in the investigation of research question 3, negotiation does show its usefulness by increasing the number of incidental exposures learners get to the vocabulary. Higher numbers of correct answers were found for the negotiated one-word verbs on both the multiple choice and sentence writing measures. That is, although gains in surface level recognition appear to be equal (as outlined in the findings reported for research question 1), form+meaning and grammatical and syntactical knowledge seemed to be more easily acquired for the one-word items as opposed to the potentially more difficult opaque, non-literal phrasal verbs. Thus the added exposures that learners received in the negotiation appear to have benefitted acquisition of the less complex forms. Further research on the interaction of incidental written exposure and targeted discussions using the incidentally exposed vocabulary is needed to help further illuminate the impact on vocabulary acquisition of different discussion formats which require target item use, such as the story retelling in Joe (1998) versus comprehension questions pointing at target words in the text. In the reported study only one question was asked at the end of each story, relating to how the mystery was solved by the detective character or who was the perpetrator, which students used for an unstructured discussion; perhaps using more structured discussion questions or comprehension questions would have resulted in greater gains.

The phrasal verbs in this study were not acquired evenly across the board which opens up the question of *why*? Some target phrasal verbs, like ***swill down*** appeared very difficult for students to acquire, as indicated by the number of incorrect answers, while others, like ***suss out***, were not. One possibility is that some written contexts were more conducive to understanding a phrasal verb, regardless of the fact that each target was

presented at least once with an appositive explanatory phrase that offered a definition. A more detailed analysis of the helpfulness of semantic contexts as well as the syntactic presentation of the targets in the mystery stories is needed to help shed light on this. It is interesting that only three phrasal verbs out of the original 16 target items were marked as known on the self-rated pre-test. The phrasal verbs were carefully chosen to not be listed in the top 100 phrasal verbs in the BNC (Gardner & Davies, 2007) and as opaque, non-literal phrasal verbs; yet, three of these items were marked as known, which was borne out by the numbers of correct form+meaning matches for these items on the multiple choice test. However, none of the one-word verbs chosen from the 9,000 frequency level of English were recognized, showing that words at this level were sufficiently infrequent for intermediate to low-advanced learners to not have already known them. Ideally, a better way of identifying infrequent phrasal verbs would have been useful; thus, the thesis study supports Gardner and Davies' (2007) call for corpus research around the frequency levels of phrasal verbs in the English language. Without more detailed corpus analysis it is impossible to predict if the phrasal verbs recognized in this thesis study belong to a more frequent lexical range than the one-word items or if another factor was at play. Other than Gardner and Davies' (2007) list of the 100 most frequent phrasal verbs, teachers must cherry-pick phrasal verbs by intuition from resources like dictionaries or trust material writers to use reliable methods used for choosing the phrasal verbs used in textbooks. In addition to delineating frequency levels of phrasal verbs, another segmentation could be done for different varieties of English such as American, British or Australian, in order to see if phrasal verbs are used as frequently among them as well as which phrasal verbs are used. This is an exciting area for future corpus research.

Implications for Instruction

The investigation of research question 1 gave insight into the numbers of exposure needed for incidental acquisition of both one-word and phrasal verbs to occur. Eight text exposures plus opportunities for oral interaction appear to be useful. The time consuming process required to build in the eight incidental textual exposures used in this study is not feasible to expect teachers to undertake. The reported gains were small compared to the amount of effort required to make the material. Also, maintaining the 98% known word threshold entails that only a few unknown words can be presented per page. This means that many pages of text would be required for a medium sized vocabulary list of 15 items to be incidentally encountered eight repeated times in unique contexts. It is also important to remember that all of the words were not learned. Since about 50% of the meanings were acquired in this study and about 40% of targets were acquired with productive level knowledge, is it justifiable to take the potentially large amount of time required to build targets into texts for a gain of only a few words? Probably not. Teachers can and do build vocabulary into texts and may also use other strategies for recycling words, but it is not realistic to expect them to engage in extensive text modifications for all the vocabulary they wish to teach considering teachers also have a number of other duties to attend to for their classes.

Using several texts revolving around the same theme may help ensure the same vocabulary is met while reading since it is likely the same semantic set is used to discuss the topic. Theme-based reading is also more likely to contain open theme-based semantic sets, which are more readily learned than closed semantic sets (Waring, 1997). This

technique is useful for both teachers and students. An easily accessible example is reading follow-up stories to newspaper articles.

Textbook materials are often designed around a core set of vocabulary that is based upon a unit's theme. In this case it would be possible for the material developer to ensure the core set of vocabulary is recycled at least eight times in unique contexts throughout each unit and preferably throughout the entire text. While lower-level texts often have shorter units, it would still be possible to have the same words appear in later units as part of the unit body or in review materials. Since students may or may not use the review material, it is advisable to ensure recycling of vocabulary occurs in the main content of each unit. The literature review outlined how graded readers recycle vocabulary among levels (Nation & Wang, 1999). With careful planning, such a system could be implemented within a textbook series.

The assumption that phrasal verbs are very difficult to acquire was negated by the surprising finding that gains for PVs and single word verbs were almost equal (research question 2). This assumption has led to the creation of courses and materials specifically geared towards learning PVs. If PVs are not as difficult to learn as assumed, are such courses and materials necessary? Is it possible for teachers to simply include a selection of the 100 most frequent PVs identified by Gardner and Davies (2007), which learners can be exposed to using incidental and instructed means? Since these 100 PVs make up half of the PV occurrences in the BNC, teachers, learners and material writers can confidently use this list as a benchmark for selecting items to teach or learn until further corpus analysis is done to provide frequency lists of English PVs. Pedagogically, the finding that PVs were acquired through reading plus oral negotiation at almost the same

rate as the single word verbs suggests that teachers can use the same methods for teaching PVs as other vocabulary items.

The usefulness of negotiation for vocabulary acquisition (outlined in the findings for research question 3) helps teachers who want to include such common communicative activities with the aim of aiding or focusing on vocabulary acquisition. For teachers using general negotiation tasks, planning an activity to include new or previously met target vocabulary is a useful way to increase learners' exposure while not having vocabulary as the focus of the activity. The potential incorrect learning from negotiation does present a caution sign for teachers. Clearly, they should to ensure that their learners have correctly matched the form+meaning before, after or during such exercises. In order to help determine that learners have correctly understood the words focused on through negotiation, teachers can employ techniques such as a quick review game of matching items and definitions or simply supplying a definition for a given item. Learners could also be required to add assigned words to a word journal which would later be checked by the teacher. If the material provides a definition, the teacher could directly ask students to find this in the text as part of a written or oral comprehension task.

In summary, this chapter has presented a number of possible avenues for future research as well as practical classroom applications. Some ideas suggested for future research included possible replications of this study by using longer texts over a greater time span and the comparison of different grammatical types of FSs. Investigating whether PVs are acquired in stages is also possible as is expanding the PV corpus started by Gardner and Davies (2007) to include more delineations of English PV frequency. Pedagogical recommendations include recycling vocabulary to increase learner

encounters. Caution was given to teachers when discussions are used as part of classroom exercises for vocabulary learning, whether the vocabulary exposure is intended or incidental. Ensuring students have made the correct form+meaning match is necessary as inaccurate learning can result from negotiation of vocabulary items. The discussion will now turn to concluding comments in the next section.

Chapter 6: Conclusion

This thesis study compared the incidental acquisition through reading of phrasal verbs, a type of formulaic sequence, with one-word verbs, thereby extending previous research which has only included phrasal verbs as part of a mixed lexical set. It also provided new insights comparing the acquisition of multi-word formulaic sequences to one-word items. The narrowly focused list of verbs used in this study has helped to shed some light on whether items within the same lexical class, although of different form, are equally acquired by the same instructional treatment. The findings reported above show that one-word verbs and phrasal verbs were acquired almost equally with an exposure rate of eight unique occurrences in the written texts coupled with negotiation. Thus, phrasal verbs appear to be more easily acquired than previously thought. Considering the difficulties many students attach to learning phrasal verbs mentioned in the introduction, it is surprising to discover phrasal verbs were almost as easily acquired as one-word verbs in this study. This relative ease of acquisition is positive news for learners, however. It was expected that negotiation would assist acquisition, and this was indeed found, but the finding that negotiation could also predict inaccurate learning was unexpected; this is useful for classroom design techniques to help ensure learners have indeed acquired the correct meaning.

Several possible areas for future research, for example expanding this study to include more participants over a longer period of time and continuing with corpus research to determine frequency levels of phrasal verbs in English, were identified. Overall, until this study can be repeated with a larger sample of students the findings need to be treated with caution. However, it is exciting that phrasal verbs may be more

accessible than previously thought. Also, it is encouraging to find tentative evidence that coupling techniques such as incidental exposure through reading with negotiation may lead to significant gains of form+meaning learning as well as acquisition of deeper syntactic and grammatical knowledge.

References

Adolphs, S., & Durow, V. (2004). Social-cultural integration and the development of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 107-126). Philadelphia: John Benjamins Publishing Company.

Bensoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading, 7*(1), 15-32.

Bishop, H. (2004). The effect of typographic salience on the look up and comprehension of unknown formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 227-248). Philadelphia: John Benjamins Publishing Company.

Bruton, A. (2009). The vocabulary knowledge scale: A critical analysis. *Language Assessment Quarterly, 6*(4), 288-297.

Coady, J. (1997). L2 vocabulary acquisition through extensive reading. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 225-237). Cambridge: Cambridge University Press.

Coady, J., & Huckin, T. (1997). *Second language vocabulary acquisition: A rationale for pedagogy*. New York: Cambridge University Press.

Coady, J., Magoto, J., Hubbard, P., Graney, J., & Mokhtari, K. (1993). High frequency vocabulary and reading proficiency in ESL readers. In T. Huckin, M. Haynes & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 217-228). Norwood, NJ: Ablex Publishing Corporation.

Cobb, T. (2009). The compleat lexical tutor (version 6.2). [Computer software]. Retrieved from http://www.lextutor.ca/

Collins COBUILD dictionary of phrasal verbs. (2002). Glasgow: HarperCollins Publishers.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238.

Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs-a case for contrastive analysis. *Studies in Second Language Acquisition, 7*, 73-80.

Day, R., Omura, C. & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language, 7*, 541-551.

de la Fuente, M. (2002). Negotiation and oral acquisition of L2 vocabulary. *Studies in Second Language Acquisition, 24*(1), 81-112.

Elley, W. (1991). Acquiring literacy in a second language: The effect of book-based programs. *Language Learning, 41*(3), 375-411.

Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly, 41*(2), 339-359.

Grabe, W., & Stoller, F. (1997). Reading and vocabulary development in a second language: A case study. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition* (pp. 98-122). Cambridge: Cambridge University Press.

Hafiz, F.M., & Tudor, I. (1989). Extensive reading and the development of language skills. *ELT Journal, 43*(1), 4-13.

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Modern Language Review, 61*(3), p. 355-382.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a *Clockwork Orange*: Acquiring second language vocabulary through reading. *Reading in a Foreign Language, 11*(2), 207-223.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *Studies in Second Language Acquisition, 21*, 181-193.

Hulstijn, J. (1992.) Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 113-125). London: MacMillan Academic & Professional Ltd.

Hulstijn, J. (2003). Incidental and intentional learning. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 349-381). Malden, MA: Blackwell Publishing Ltd.

Hulstijn, J., & Laufer, B. (2001). Some empirical evidence for the Involvement Load hypothesis in vocabulary acquisition. *Language Learning, 51*(3), 539-558.

Hulstijn, J., & Marchena, E. (1989). Avoidance: Grammatical or semantic causes? *Studies in Second Language Acquisition, 11*, 241-255.

Irujo, S. (1986a). Don't put your leg in your mouth: Transfer in the acquisition of idioms in the second language. *TESOL Quarterly, 20*(2), 287-304.

Irujo, S. (1986b). A piece of cake: learning and teaching idioms. *ELT Journal, 40*(3), 236-242.

Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal, 88*(3), 416-432.

Joe, A. (1995). Text-based tasks and incidental vocabulary learning. *Second Language Research, 11*, 149–158.

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics, 19*(3), 357-377.

Kellerman, E. (2000). What fruit can tell us about lexicosemantic transfer: A non-structural dimension to learner's perceptions of linguistic relations. In C. Munoz (Ed.), *Segundas lenguas adquisicion en la aula* (pp. 21-39). Barcelona: Ariel.

Kilgareff, A., & Rundell, M. (2006). Efficient corpus development for lexicography: building the New Corpus for Ireland. *Language Resources and Evaluation Journal, 40*(2), 127-152.

Krashen, S. (1985). The input hypothesis: Issues and implications. New York: Longman.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal, 73*(4), 440-464.

Krashen, S. (1991). The input hypothesis: An update. In J.E. Alatis (Ed.), *Georgetown University round table on languages and linguistics,1991* (pp. 409-431). Washington, DC: Georgetown University Press.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. Arnaud & H. Bejoint (Eds.), *Vocabulary and Applied Linguistics* (pp. 126-132). London: MacMillan Academic and Professional, Ltd.

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 20–34). New York: Cambridge University Press.

Laufer, B. (2006). Comparing Focus on Form and Focus on FormS in second-language vocabulary learning. *Canadian Modern Language Review, 63* (1), 149-166.

Laufer, B., & Eliasson, S. (1993). What causes avoidance in L2 learning? L1-L2 difference, L1-L2 similarity, or L2 complexity? *Studies in Second Language Acquisition, 15*, 35-48.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: size, strength, and computer adaptiveness. *Language Learning, 54*(3), 399-436.

Laufer, B. & Ravenhorst-Kalovski, G. (2010). Lexical threshold revisited: Lexical text coverage, learner's vocabulary size and reading comprehension. *Reading in a Foreign Language, 22* (1), 15-30.

Lewis, M. (1997). Pedagogical implications of the lexical approach. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 255–270). New York: Cambridge University Press.

Lightbown, P. (1992). Can they do it themselves? A comprehension-based ESL course for young children. In R. Courchene, J. St. John, E. Therien, & J. Glidden (Eds.), *Comprehension-based language teaching: Current trends* (pp. 353-370). Ottawa: University of Ottawa Press.

Lightbown, P. (2000). Anniversary article: Classroom SLA research and second language teaching. *Applied Linguistics, 21*(4), 431-462.

Lightbown, P. (2003). SLA research in the classroom/SLA research for the classroom. *Language Learning Journal, 28*, 4-13.

Lightbown, P. (2008). Transfer appropriate processing as a model for second language acquisition. In Z. Han (Ed.), *Understanding second language processes* (pp. 27-44). Clevedon: Multilingual Matters.

Lightbown, P., Halter, R., White, J., & Horst, M. (2002). Comprehension-based learning: The limits of do-it-yourself. *Canadian Modern Language Review, 58*(3), 427-464.

Long, M. (1984). Process and product in ESL program evaluation. *TESOL Quarterly, 18*(3), 409-425.

Long, M. (1996). The role of linguistic environment in second language acquisition. In W.C. Ritchie & T.K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). San Diego, CA: Academic Press.

Loschky, L. (1994). Comprehensible input and second language acquisition: What is the relationship? *Studies in Second Language Acquisition (SSLA), 16*, 303-323.

Nagy, W., Herman, P., & Anderson, R. (1985). Learning words from context. *Reading Research Quarterly, 20*(2), 233-253.

Nassaji, H. (2004). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Canadian Modern Language Review, 61*(1), 107-134.

Nation, I. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation. I. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*(1), 59-82.

Nation, I., & Wang Ming-tzu, K. (1999). Graded readers and vocabulary. *Reading in a Foreign Language, 12*(2), 355-380.

Paribakht, T.S., & Wesche, M. (1993). The relationship between reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal, 11*(1), 9-29.

Paribakht, T. S., & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 174–199). New York: Cambridge University Press.

Parry, K. (1997). Vocabulary and comprehension: Two portraits. In J. Coady & T. Huckin (Eds.), *Second language vocabulary* (pp. 55-68). Cambridge: Cambridge University Press.

Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J.C. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 191-226). London: Longman.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language, 18*(1), 1-28.

Pitts, M., White, H., & Krashen, S. (1989). Acquiring a second language vocabulary through reading: A replication of the *Clockwork Orange* study using second language acquirers. *Reading in a Foreign Language, 5*(2), 271-275.

Prodromou, L. (2003). Idiomaticity and the non-native speaker. *English Today, 74*(19-2), 42-48.

Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review, 56*(2), 282-308.

Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*(3), 513-536.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing,10*, 355-371.

Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A.J. Kunnan (Ed.), *Validation in language assessment: Selected papers from the 17th language testing research colloquium* (pp. 41-60). Mahwah, NJ: L. Erlbaum.

Read, J., & Nation, P. (2004). Measurement of formulaic sequences. In N. Schmitt (Ed.), *Formulaic Sequences* (pp.23-35). Philadelphia: John Benjamins Publishing Company.

Saragi, T., Nation, P., Meister, G. (1978). Vocabulary learning and reading. *System, 6*(2), 72-80.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11,* 17-45.

Schmidt, R. (1993). Awareness and second language acquisition. In W. Grabe (Ed.), *Annual review of applied linguistics, 13: Issues in second language teaching and learning* (pp. 206–226). New York: Cambridge University Press.

Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. In J.H. Hulstijn & R. Schmidt (Eds.), *Consciousness in second language learning*, *AILA Review,* 11 (Special issue), 11-26.

Schmitt, N., Dornyei, Z., Adolphs, S., & Durow, V. (2004). Knowledge and acquisition of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 107-126). Philadelphia: John Benjamins Publishing Company.

Schmitt, N., Jiang, X., Grabe, W. (March, 2008). *The percentage of words known in a text and reading comprehension*. Paper presented at the American Association of Applied Linguistics (AAAL), Washington, DC.

Sobol, D. (1967). *Two-minute mysteries*. New York: Scholastic Inc.

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *The Canadian Modern Language Review*, *50(*1), 158-164.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics* (pp. 125-144). Oxford: Oxford University Press.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics, 16*, 371-391.

Verspoor, M., & Lowie, W. (2003). Making sense of polysemous words. *Language Learning, 53*(3), 547-586.

Waring, R., & Nation, I. (2004). Second language reading and incidental vocabulary learning. *Angles on the English Speaking World, 4*, 11-23.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*(2), 130-163.

Wode, H. (1999). Incidental vocabulary acquisition in the foreign language classroom. *Studies in Second Language Acquisition, 21*, 243-258.

Wolter, B. (2001). Comparing the L1 and L2 mental lexicon. *Studies in Second Language Acquisition, 23,* 41-69.

Wray, A. (2002). Formulaic language and the lexicon. Cambridge: Cambridge University Press.

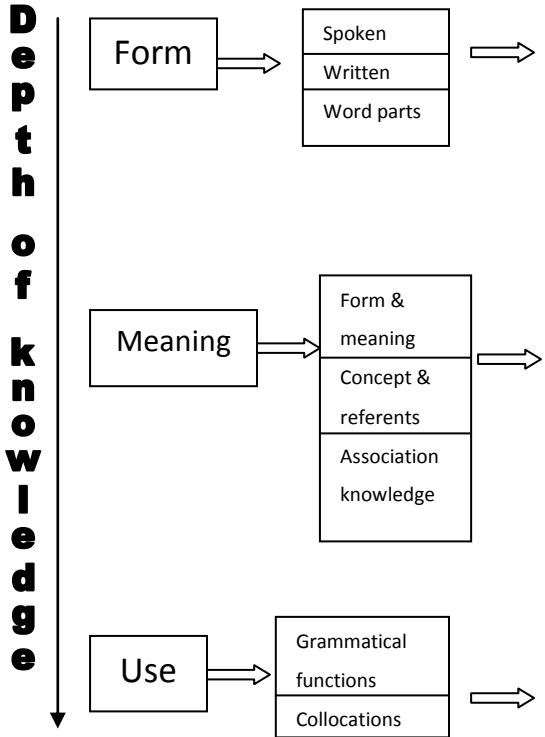
Wray, A., & Perkins, M. (2000). The functions of formulaic language: an integrated model. *Language & Communication*, *20*, 1-28.

Yorio, C. (1989). Idiomaticity as an indicator of second language proficiency. In K. Hyltenstam & L. Obler (Eds.), *Bilingualism across the lifespan* (pp.55-72). Cambridge: Cambridge University Press.

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *The Canadian Modern Language Review, 57*(4), 541-572.

Appendix A

Nine levels of word knowledge (Nation, 2001)



Spoken	Receptive: sound
	Productive: pronunciation
Written	Receptive: word's appearance
	Productive: spelling and orthography
Word parts	Receptive: recognizable parts
	Productive: parts necessary to express meaning

Form & Meaning	Receptive: meaning signaled by word form
	Productive: word form to express meaning
Concept & Referents	Receptive: concept information details
	Productive: possible items concept can point to
Association Knowledge	Receptive: related semantic set activated by word
	Productive: possible synonyms

Grammatical function	Receptive: possible word patterns
	Productive: restricted word patterns of use
Collocations	Receptive: possible additional words/ word types used
	Productive: necessary additional words/word types used
Constraints on use	Receptive: expected frequency and environment of encounter
	Productive: actual frequency and environment of encounter

Appendix B

SELF-RATED VOCABULARY TEST

NAME: _____

DATE:_____

DIRECTIONS: For each word, please check (✓) YES if you know the word and the meaning, NOT SURE if you might know the word and the meaning, or NO if you do not know the word.

E.g. happiness _✓__ YES ___NOT SURE ___NO

1. pipe	___YES ___NOT SURE __NO	20. calm	___YES ___NOT SURE ___NO
2. dump	___YES ___NOT SURE ___NO	21. iron	___YES ___NOT SURE ___NO
3. seek	___YES ___NOT SURE ___NO	22. update	___YES ___NOT SURE ___NO
4. fail	___YES ___NOT SURE ___NO	23. bring in	___YES ___NOT SURE ___NO
5. wakeful	___YES ___NOT SURE ___NO	24. plug	___YES ___NOT SURE ___NO
6. hold out	___YES ___NOT SURE ___NO	25. lift	___YES ___NOT SURE ___NO
7. guard	___YES ___NOT SURE ___NO	26. smelly	___YES ___NOT SURE ___NO
8. fuel	___YES ___NOT SURE ___NO	27. racetrack	___YES ___NOT SURE ___NO
9. object	___YES ___NOT SURE ___NO	28. grade	___YES ___NOT SURE ___NO
10. draw up	___YES ___NOT SURE ___NO	29. give up	___YES ___NOT SURE ___NO
11. demonstrator	___YES ___NOT SURE ___NO	30. hunger	___YES ___NOT SURE ___NO
12. extremely	___YES ___NOT SURE ___NO	31. split	___YES ___NOT SURE ___NO
13. huge	___YES ___NOT SURE ___NO	32. pelt	___YES ___NOT SURE ___NO
14. net	___YES ___NOT SURE ___NO	33. warn	___YES ___NOT SURE ___NO
15. persuasive	___YES ___NOT SURE ___NO	34. dance	___YES ___NOT SURE ___NO
16. landlord	___YES ___NOT SURE ___NO	35. queenly	___YES ___NOT SURE ___NO
17. generated	___YES ___NOT SURE ___NO	36. occupied	___YES ___NOT SURE ___NO
18. handle	___YES ___NOT SURE ___NO	37. turn out	___YES ___NOT SURE ___NO
19. set off	___YES ___NOT SURE ___NO	38. valid	___YES ___NOT SURE ___NO

39. thick	___YES ___NOT SURE ___NO
40. acid	___YES ___NOT SURE ___NO
41. look around	___YES ___NOT SURE ___NO
42. put down	___YES ___NOT SURE ___NO
43. maintain	___YES ___NOT SURE ___NO
44. theory	___YES ___NOT SURE ___NO
45. mill	___YES ___NOT SURE ___NO
46. metallic	___YES ___NOT SURE ___NO
47. get back	___YES ___NOT SURE ___NO
48. float	___YES ___NOT SURE ___NO
49. get up	___YES ___NOT SURE ___NO
50. spell out	___YES ___NOT SURE ___NO
51. vary	___YES ___NOT SURE ___NO
52. weird	___YES ___NOT SURE ___NO
53. owe	___YES ___NOT SURE ___NO
54. come back	___YES ___NOT SURE ___NO
55. suss out	___YES ___NOT SURE ___NO
56. look up	___YES ___NOT SURE ___NO
57. revise	___YES ___NOT SURE ___NO
58. impact	___YES ___NOT SURE ___NO
59. set out	___YES ___NOT SURE ___NO
60. get out	___YES ___NOT SURE ___NO
61. lengthy	___YES ___NOT SURE ___NO
62. combine	___YES ___NOT SURE ___NO
63. honour	___YES ___NOT SURE ___NO
64. temperature	___YES ___NOT SURE ___NO
65. planter	___YES ___NOT SURE ___NO
66. turn up	___YES ___NOT SURE ___NO

67. nab	___YES ___NOT SURE ___NO
68. narrow	___YES ___NOT SURE ___NO
69. move on	___YES ___NOT SURE ___NO
70. ultimate	___YES ___NOT SURE ___NO
71. advise	___YES ___NOT SURE ___NO
72. juicy	___YES ___NOT SURE ___NO
73. hold up	___YES ___NOT SURE ___NO
74. set up	___YES ___NOT SURE ___NO
75. come out	___YES ___NOT SURE ___NO
76. break down	___YES ___NOT SURE ___NO
77. dozen	___YES ___NOT SURE ___NO
78. salt away	___YES ___NOT SURE ___NO
79. observe	___YES ___NOT SURE ___NO
80. carry out	___YES ___NOT SURE ___NO
81. witness	___YES ___NOT SURE ___NO
82. icily	___YES ___NOT SURE ___NO
83. appealing	___YES ___NOT SURE ___NO
84. forgiven	___YES ___NOT SURE ___NO
85. give in	___YES ___NOT SURE ___NO
86. take on	___YES ___NOT SURE ___NO
87. buck up	___YES ___NOT SURE ___NO
88. ordinary	___YES ___NOT SURE ___NO
89. refusal	___YES ___NOT SURE ___NO
90. hide	___YES ___NOT SURE ___NO
91. jointed	___YES ___NOT SURE ___NO
92. justice	___YES ___NOT SURE ___NO
93. feature	___YES ___NOT SURE ___NO
94. harm	___YES ___NOT SURE ___NO

95. withdrawal	___YES ___NOT SURE ___NO
96. bring up	___YES ___NOT SURE ___NO
97. wisdom	___YES ___NOT SURE ___NO
98. jacket	___YES ___NOT SURE ___NO
99. bang	___YES ___NOT SURE ___NO
100. go back	___YES ___NOT SURE ___NO
101. holy	___YES ___NOT SURE ___NO
102. pack in	___YES ___NOT SURE ___NO
103. sit back	___YES ___NOT SURE ___NO
104. establish	___YES ___NOT SURE ___NO
105. tune	___YES ___NOT SURE ___NO
106. take up	___YES ___NOT SURE ___NO
107. bicker	___YES ___NOT SURE ___NO
108. justify	___YES ___NOT SURE ___NO
109. scared	___YES ___NOT SURE ___NO
110. swill down	___YES ___NOT SURE ___NO
111. yard	___YES ___NOT SURE ___NO
112. kick	___YES ___NOT SURE ___NO
113. neighbour	___YES ___NOT SURE ___NO
114. invite	___YES ___NOT SURE ___NO
115. qualify	___YES ___NOT SURE ___NO
116. variably	___YES ___NOT SURE ___NO
117. lumpy	___YES ___NOT SURE ___NO
118. come up	___YES ___NOT SURE ___NO
119. locate	___YES ___NOT SURE ___NO
120. sit down	___YES ___NOT SURE ___NO
121. tidy	___YES ___NOT SURE ___NO
122. mortgage	___YES ___NOT SURE ___NO

123. roof	___YES ___NOT SURE ___NO
124. taste	___YES ___NOT SURE ___NO
125. nerve	___YES ___NOT SURE ___NO
126. put up	___YES ___NOT SURE ___NO
127. madden	___YES ___NOT SURE ___NO
128. joke	___YES ___NOT SURE ___NO
129. find out	___YES ___NOT SURE ___NO
130. element	___YES ___NOT SURE ___NO
131. flaunt	___YES ___NOT SURE ___NO
132. bring back	___YES ___NOT SURE ___NO
133. offence	___YES ___NOT SURE ___NO
134. annoyed	___YES ___NOT SURE ___NO
135. mucky	___YES ___NOT SURE ___NO
136. go out	___YES ___NOT SURE ___NO
137. yellow	___YES ___NOT SURE ___NO
138. nonseasonal	___YES ___NOT SURE ___NO
139. voice	___YES ___NOT SURE ___NO
140. behaviour	___YES ___NOT SURE ___NO
141. knee	___YES ___NOT SURE ___NO
142. work out	___YES ___NOT SURE ___NO
143. injured	___YES ___NOT SURE ___NO
144. decent	___YES ___NOT SURE ___NO
145. go down	___YES ___NOT SURE ___NO
146. patch	___YES ___NOT SURE ___NO
147. piglet	___YES ___NOT SURE ___NO
148. keenest	___YES ___NOT SURE ___NO
149. surface	___YES ___NOT SURE ___NO
150. fleck	___YES ___NOT SURE ___NO

151. boil	___YES ___NOT SURE ___NO
152. rich	___YES ___NOT SURE ___NO
153. ignorance	___YES ___NOT SURE ___NO
154. hilly	___YES ___NOT SURE ___NO
155. upset	___YES ___NOT SURE ___NO
156. enable	___YES ___NOT SURE ___NO
157. react	___YES ___NOT SURE ___NO
158. flinch	___YES ___NOT SURE ___NO
159. method	___YES ___NOT SURE ___NO
160. guilty	___YES ___NOT SURE ___NO
161. citizen	___YES ___NOT SURE ___NO
162. padded	___YES ___NOT SURE ___NO
163. monitor	___YES ___NOT SURE ___NO
164. typical	___YES ___NOT SURE ___NO
165. chiefly	___YES ___NOT SURE ___NO
166. rocky	___YES ___NOT SURE ___NO
167. fancy	___YES ___NOT SURE ___NO
168. take over	___YES ___NOT SURE ___NO
169. make out	___YES ___NOT SURE ___NO
170. stifle	___YES ___NOT SURE ___NO
171. reflective	___YES ___NOT SURE ___NO
172. verse	___YES ___NOT SURE ___NO
173. stylish	___YES ___NOT SURE ___NO
174. gap	___YES ___NOT SURE ___NO
175. pick up	___YES ___NOT SURE ___NO
176. crack	___YES ___NOT SURE ___NO
177. tank	___YES ___NOT SURE ___NO
178. bakery	___YES ___NOT SURE ___NO

179. youthfully	___YES ___NOT SURE ___NO
180. scrawl	___YES ___NOT SURE ___NO
181. make up	___YES ___NOT SURE ___NO
182. stone	___YES ___NOT SURE ___NO
183. fritter away	___YES ___NOT SURE ___NO
184. lend	___YES ___NOT SURE ___NO
185. edge	___YES ___NOT SURE ___NO
186. carry on	___YES ___NOT SURE ___NO
187. sack	___YES ___NOT SURE ___NO
188. psychologically	___YES ___NOT SURE ___NO
189. victim	___YES ___NOT SURE ___NO
190. eventful	___YES ___NOT SURE ___NO
191. boniest	___YES ___NOT SURE ___NO
192. move in	___YES ___NOT SURE ___NO
193. bitten	___YES ___NOT SURE ___NO
194. dirty	___YES ___NOT SURE ___NO
195. grab	___YES ___NOT SURE ___NO
196. abuser	___YES ___NOT SURE ___NO
197. break up	___YES ___NOT SURE ___NO
198. point out	___YES ___NOT SURE ___NO
199. analyse	___YES ___NOT SURE ___NO
200. quotation	___YES ___NOT SURE ___NO

Appendix C

NOTE: The original test was in Landscape page layout but has been adapted to Portrait layout for inclusion in the appendices.

NOTE: The question sheet below includes the answers.

DIRECTIONS: Match the items on the left with the meanings on the right.

Here is an example:

1. agree with
2. dictionary ___say something, mention a new thing
3. bring up ___a book with words and meanings
4. stereo ___find out where something comes from
5. be born with
6. trace back to

You answer it in the following way:

1. agree with
2. dictionary _3__say something, mention a new thing
3. bring up _2__a book with words and meanings
4. stereo _6__find out where something comes from
5. be born with
6. trace back to

You do not have to find a meaning for every word. In the example above, the words *agree with, stereo, be born with* do not have a meaning.

Try to do every part of the test. Good luck!

#1	1. turn around 2. pelt 3. move up 4. access 5. hire 6. fritter away	_5__give someone a job/work _2__fall or hit with very heavily _6__waste a resource (e.g. money) by spending it in a foolish way		#6	1. pat 2. flinch 3. swill down 4. gather 5. give in 6. move back	_3__drink fast _4__bring many things together _2__make a slight move backwards because of something frightening
#2	1. stifle 2. give back 3. salt away 4. manufacture 5. set down 6. wrap	_3__save money _2__return something _1__stop something from happening		#7	1. set about 2. distribute 3. turn off 4. fleck 5. spell out 6. volunteer	_4__a tiny mark on something _5__explain in detail or in a very clear way _1__start doing something in a purposeful way
#3	1. draw up 2. grow up 3. bicker 4. obtain 5. survive 6. take in	_4__get something _1__prepare a document _3__fight about unimportant things		#8	1. buck up 2. lack 3. flaunt 4. come about 5. unwrap 6. put off	_6__delay an event _1__make a person feel more cheerful _3__show something you are proud of so other people will admire it
#4	1. chase 2. come off 3. give out 4. justify 5. pack in 6. scrawl	_6__write in an untidy and careless way _4__give or be a good reason for something _5__stop having a romantic relationship with someone		#9	1. forgive 2. initiate 3. hold back 4. quote 5. pick out 6. turn over	_2__start an activity _4__say the cost of something _5__choose something from a group
#5	1. escape 2. suss out 3. break off 4. tour 5. nab 6. get through	_2__solve a puzzle _5__arrest, seize or grab _3__separate part of something by force		#10	1. sit up 2. bring down 3. nurse 4. rob 5. bend 6. come round	_1__move into an upright sitting position _3__take care of someone who is sick or injured _4__take something from a person or place illegally

Appendix D

DIRECTIONS:
Give your own opinion. Answer using the word in brackets ( ).
Write a complete sentence.

Example: What is the next step after making a plan? (carry)
Answer: I need to carry out the plan.

1. If you suddenly see something scary or frightening, what do you do? (flinch)

_____
_____
_____

2. Do you know anyone who ended a relationship after a long time? (pack)

_____
_____
_____

3. Where is the best place to keep money? (salt)

_____
_____
_____

4. What month does it rain the hardest? (pelt)

_____
_____
_____

5. If the class is taking a test and suddenly you need to laugh out loud, what do you do? (stifle)

_____
_____
_____

6. What are police detectives good at? (suss)

_____
_____
_____

7. What does a lawyer do? (draw)

_____
_____
_____

8. Do you always like to be busy? (fritter)

_____
_____
_____

9. How would people describe your handwriting? (scrawl)

_____
_____
_____

10. Briefly describe your relationship with your best friend. (bicker)

_____
_____
_____

11. What do some people do with their favourite drink? (swill)

_____
_____
_____

12. Your friend calls you and is upset, what do you do? (buck)

_____
_____
_____

13. What would you do if you won two million dollars? (flaunt)

_____
_____
_____

14. What do police officers want to do with criminals? (nab)

_____
_____
_____

15. When your father or mother was fifty-five years old, what colour was his/her hair? (fleck)

_____
_____
_____

16. Do you need a recipe when you cook food? (spell)

_____
_____
_____