Incidental Vocabulary Acquisition through Aural Means:

What Do English Television Programs Have to Offer?

Myriam Lévesque

A Thesis

in

The Department

of

Education

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Arts (Applied Linguistics) at

Concordia University

Montreal, Quebec, Canada

March 2013

Concordia University

School of Graduate Studies

This is to certify that the thesis prepared

By: Myriam Lévesque

Entitled: Incidental Vocabulary Acquisition through Aural Means: What Do
              English Television Programs Have to Offer?

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
Dr. Joanna White

_____ Examiner
Dr. Marlise Horst

_____ Examiner
Dr. Pavel Trofimovich

_____ Supervisor
Dr. Walcir Cardoso

Approved by _____
            Chair of Department or Graduate Program Director

_____ 2013   _____
                            Dean of Faculty

Abstract

Incidental Vocabulary Acquisition through Aural Means:
What Do English Television Programs Have to Offer?

Myriam Lévesque

Little research has been done with regards to the lexical environment provided by broadcasting television. In this study, the contribution of television viewing to incidental vocabulary acquisition was explored using a variety of corpora. Sitcoms, dramas and sci-fi programs were compared among themselves and to a teacher talk corpus to determine how television programs can benefit learners' acquisition of vocabulary, particularly when combined with teacher talk. It was shown that the learnable vocabulary of teacher talk and TV-talk do not overlap. Even though dramas are slightly easier to understand than sitcoms and sci-fi programs, all three cater to the needs of advanced English as a second language students. Sci-fi programs appear to offer the most learnable words of all the programs. Because the lexical load of teacher talk is less demanding than that of TV-talk, combining the two activities to increase the number of learnable words available proves unfeasible. Implications for incidental vocabulary learning, research and pedagogy are discussed.

Acknowledgements

Table of Content

List of Figures

List of Tables

Chapter 1 - Introduction

Many teachers, including colleagues and myself, give "television homework" to students hoping that the additional exposure to the second language (L2) will eventually yield vocabulary acquisition. The rationale behind this idea is manifold. First, it is now well known that repeated exposure to vocabulary items is necessary for acquisition to take place (e.g., Elley, 1989; Huckin & Coady, 1999; Krashen, 1985; Meara, 1993; Pearson, Fernandez, Lewedeg & Oller, 1997). Secondly, this activity allows students to practice their aural comprehension skills. The situation in Quebec is such that the budgetary constraints often make it difficult for many teachers to expose learners to native or fluent speakers to develop these skills. Thus, TV homework has become a means to address these limitations. Finally, it has been shown that children, teenagers and adults alike engage in television watching as leisure activity on a regular basis (Jopp & Hertzog, 2010; Nippold, Duthie & Larsen, 2005). Therefore, there is a potential for teachers to combine students' interests and the possible benefits of TV watching, thus providing activities that offer aural comprehension practice as well as vocabulary acquisition opportunities.

Some of the questions that arise from such practice are whether television programs are in fact environments with rich vocabulary acquisition potential, if and what genre of TV show is better indicated for learners of a particular proficiency level, and whether the combination of TV homework and teacher talk is as beneficial with regards to vocabulary acquisition as teachers believe. Along these lines, the goal of this corpora-based research is to identify the benefits of adding 50 hours of television viewing to an English as a Second Language (ESL) language course.

It has been shown that some vocabulary acquisition from television exposure is possible, but most of the research has been concerned with the types of support (e.g. subtitles, captions) that yield the best result (e.g., D'Ydewalle & Van de Poel, 1999; Koolstra & Beentjes 1999; Koskinen, Knable, Markham, Jensema & Kane, 1996; Rice & Woodsmall, 1988). But, little has been done to determine the richness of the TV environment per se (exceptions are MacFadden, Barrett & Horst, 2009; Webb & Rodgers, 2009). MacFadden et al. (2009) have produced a specialized Television Word List that covers 1 to 2% of TV talk which, if learnt, could potentially boost learners understanding of drama and sitcom shows. They have identified a list of English television words that may be problematic to second language learners. Meanwhile Webb and Rogers (2009) have looked at the vocabulary coverage and number of encounters of low-frequency words in TV programs according to genres. It would be interesting to see if the results of the current research support the observations made by these researchers. Questions about what genre of programs is best suited to learners depending on their proficiency levels (using Nation's (2001) classification) and questions about how TV programs can complement teacher talk in terms of incidental vocabulary acquisition remain largely unanswered. This reality can be attributed to the fact that most of the research on vocabulary acquisition is concerned with learning through exposure to written text, not audio- or video-based input (Day, Omura & Hiramatsu, 1991; Horst, Cobb & Meara, 1998; Nagy, Anderson & Herman, 1987; Krashen, 1989). This thesis will address these gaps in the literature by studying the learnable vocabulary available in sitcoms, dramas and sci-fi programs and show how they complement the learnable vocabulary available in teacher-talk.

The remainder of this thesis is organized in the following manner. Chapter 2 provides a review of the relevant literature, which will introduce the concepts of incidental vocabulary acquisition, a type of learning that could potentially be exploited through TV watching. An overview of Krashen's (1985) $i + 1$ theory and of research on incidental vocabulary acquisition through reading precedes the discussion on incidental vocabulary acquisition through aural means. This is because these concepts are the roots of many studies in the latter field. A discussion of the conditions necessary for such acquisition to occur when exposed to aural media will support the use of television viewing for vocabulary learning. In addition, the findings of previous research will justify the thresholds used in this research – 98% known-word coverage and the number of word repetitions (10) – to identify potential learnable vocabulary. Finally, the research with regards to the lexical environment of teacher talk and TV-talk is summarized. It is shown that teacher talk suffers some limitations with regards to the availability of incidentally learnable vocabulary (Horst, 2010) and that studies about television lexical environments would benefit from using larger samples of TV-talk.

Chapter 3 follows this literature review, where I present the methodology and the analytical tools that guided my study. The British National Corpus (BNC) classification and a theme classification are used. Chapter 4 describes the results to the three research questions while Chapter 5 discusses these results in the light of the previous literature review. Finally, Chapter 6 highlights the implications of the findings with regards to pedagogical applications and possibilities for future research.

Chapter 2 - Literature review

In order to determine the kind of contribution, if any, television programs can make to learner's incidental vocabulary acquisition, we must first identify the necessary conditions for such learning occur. Krashen's theory of learning from comprehensible input is at the core of all the research on the subject. Most studies focus on incidental vocabulary acquisition through reading. Their findings about word repetitions and learner's proficiency levels, texts' density of unknown words and learners' language processing strategies needed for incidental word acquisition have been used in research about incidental acquisition through aural means. This is mainly because of the lack of knowledge about the conditions required for such learning to occur through this means. In the next section, I will examine the literature that led to the 98% known-word coverage and word repetition threshold (10) used to identify learnable vocabulary in aural texts. A closer look at the research about exposure to teacher talk and TV-talk will lead to the articulation of the current research questions.

Krashen's theory of learning from comprehensible input

According to Krashen's (1985) comprehensible input theory, learners of a second language will acquire vocabulary if the language to which they are exposed is slightly more difficult than the level at which they are comfortable. He argues that this input is essential for developing comprehension and new language knowledge. In a follow-up study, Krashen (1989) concluded that all teaching is inefficient to some degree and that learners will acquire vocabulary from reading more efficiently than they would from classroom teaching on the condition that they are exposed to texts that are just above their

reading competency level. This type of learning is known as incidental vocabulary acquisition. Hulstijn and Laufer (2001, p. 539) define it as being "the [processing] of new information without the intention to commit this information to memory". In other words, the learning of new vocabulary is a by-product of another activity. It contrasts with intentional learning in that this "activity [aims] at committing [the] lexical information to memory". The current research focuses on the incidental type of learning, which will be discussed in the next section.

Conditions for successful incidental vocabulary acquisition through reading

There are very few studies to date on incidental vocabulary acquisition through aural means. As a result, many researchers have based their investigations on the findings about vocabulary acquisition through reading.

Many of these studies support Krashen's (1985) theory, described earlier, thus showing that it is possible to learn vocabulary through reading. For example, Elley (1989) investigated the vocabulary acquisition of seven and eight-year olds through repeated oral story reading. Having heard the story three times during a week, without having received any explanation, learners showed gains of 15 percent on a pictorial and synonym test. These gains appeared to be relatively permanent as shown by the delayed follow-up test. The author concluded that the best predictors of vocabulary retention were the "frequency of the word in the text, depiction of the word in illustrations, and the amount of redundancy in the surrounding context" (Elley, 1989, p. 174).

In another investigation of incidental vocabulary acquisition through reading, Day, Omura and Hiramatsu (1991) studied high school and university Japanese students who read a short story for entertainment. They also found that incidental vocabulary

acquisition was possible, as the participants learnt on average 1 to 3 words during the study as shown by their results on the multiple-choice test that followed reading. However, given the limited scope of their research, it is difficult to determine how much permanent vocabulary knowledge can be derived from this activity. Indeed, the story consisting of approximately a thousand words was only read once, only 17 words were tested, and no post-test was used. By contrast, Dupuy and Krashen (1993) found more positive results with university learners of French as a foreign language – showing average vocabulary gains of 7 words after having watched the first 5 scenes and read the subsequent 5 scripts of *Trois hommes et un couffin (1985)*. The combination of the viewing of scenes and reading of script excerpts could potentially explain these higher results. Nevertheless, just like with the study by Day et al. (1991), very few words were actually tested – only 8 of the 30 words tested appeared in the movie – and so, it is difficult to draw generalizations from these results.

In a study of greater scope, Horst, Cobb and Meara (1998) make use of a much longer text and also take into account long-term acquisition. Their findings corroborate the results obtained in other studies: incidental vocabulary acquisition is possible. They performed a quasi-experimental study in which a group-class of 34 low-intermediate learners of English read the simplified text of *The Mayor of Casterbridge*. On a test combining multiple-choice items – where the participants had to choose the correct definition – and word-association items – where the participants had to make a meaning link between two words by eliminating the third odd one – average gains of 5 words for an exposure to 21,000 words were observed. Similar to Elley's (1989) study, the authors concluded that frequent exposure to words within a text appeared to positively impact its

acquisition, even though the gains were fairly limited. In addition to confirming that incidental vocabulary acquisition is possible, these studies show that it does require a significant amount of exposure to learn new words, as they need to be encountered frequently in order to be fully acquired.

Researchers and teachers may question the importance of incidental vocabulary acquisition altogether, as it has been shown that focused attention to words is more effective in the acquisition of new words. For example, Elley (1989) has shown gains of 40% with focused attention on vocabulary, compared to 15% with incidental acquisition. In addition, Carver (1994) warns that the method of free-reading as a vocabulary growth activity is often misguided because learners rarely read at a level higher than their own. However, research has shown that if the learners are required to do readings that are slightly above their level, vocabulary gains are observed (e.g., Dupuy & Krashen, 1993; Horst, Cobb & Meara, 1998). This implies that focused learning could be complemented by incidental acquisition if this activity is conducted with proper guidance. This idea could then potentially be extended to vocabulary repeated in the TV programs that learners already watch at home to supplement in-class learning in hopes of improving their vocabulary.

Word repetition and proficiency

Many researchers believe that the frequent reoccurrence of words in written texts is necessary for vocabulary acquisition to occur: more frequently occurring items are more likely to be acquired than those that rarely occur in the language to which the learner is exposed (e.g., Horst, Cobb & Meara, 1998; Rott, 1999; Zahar, Cobb & Spada, 2001). However, the number of required repetitions remains unclear. For example, Zahar,

Cobb and Spada (2001) summarize the literature on the subject and note that this number ranges from 6 to 20 encounters. In their research, they observed that the proficiency level of learners impacted this number by increasing the likelihood of vocabulary acquisition with increased proficiency. Similarly, Horst, Cobb and Meara (1998) observed that learners with a greater vocabulary repertoire, i.e., higher proficiency, acquired more words than those who had a more limited vocabulary. In a study involving Danish learners of English, Henriksen, Alebrechtsen and Haastrup (2004) also found that second language (L2) vocabulary size is a strong predictor of reading skills and that learners with small vocabulary size had difficulties with L2 reading.

In addition to the learners' vocabulary size, Horst et al. (1998) found that the recycling of unfamiliar words also appears to promote acquisition. In general, however, frequency appears to play the largest role with lower level learners, as exemplified by a study conducted by Rott (1999) with intermediate learners. She found that 6 or more repetitions of unknown words make a significant difference in incidental vocabulary acquisition through reading. To this effect, Horst and Meara (1999) developed matrices that predicted the learning effects of repeated encounters with words in story contexts. On this note, Zahar et al. (2001) hypothesized that higher-level learners can more easily use the context to decipher unknown words, as lower level students need more encounters to find a richer context to help them decode the meaning of novel words. Their findings, however, led the authors to the conclusion that "the effect of contextual support appears to be subordinate to frequency" (Zahar et al., 2001, p. 555). It is important to note that the very definition of "knowing a word" impacts the number of occurrence necessary to acquire the said word. Accordingly, this number is likely to vary whether "knowing a

word" means being able to recognize a definition of the said word or being able to identify contexts in which it is used correctly. For the purpose of the present study, a known word is a word for which a definition or a translation of the word can be recognized. The number of words known by a learner is important because it will determine the vocabulary available for acquisition in the various corpora. In effect, only the unknown words repeated enough times in a corpus are available for uptake; to identify these unknown words, one needs to know which words the targeted learners are likely to already know. As we will discuss, below the ratio of known words to unknown words is paramount to identify the potentially learnable words.

Density of unknown words

Knowing that learners can acquire vocabulary through reading texts that are slightly above their proficiency level, researchers have sought to determine the threshold at which learners could maximize their learning. In addition to the roles of repetition and learner's proficiency, Na and Nation (1985) found that the density of unknown words, i.e., the percentage of unknown words in a given text, also affected the difficulty of inferring the meaning of an unknown word, an essential part of the process for vocabulary learning to take place. As expected, words appear to be easier to decipher when fewer unknown words are present. So, the question that then arises is how many known words are needed to reach the density of known vocabulary necessary for accurately interpreting the meaning of an unknown word. This is the topic of the forthcoming discussion.

In the case of English, Laufer (1992) established the vocabulary threshold for reading comprehension at 3,000 word families. A word family consists of the base word

(e.g., *love*), its lemmas (e.g., *loves*, *loved*, *loving*) and its basic derived forms (e.g., *lover*, *lovers*). In other words, the density of unknown words became manageable to the learners when they knew at least 3,000 high frequency English word families. These word families offered a density of known words that allowed the accurate guessing of enough unknown words to enhance reading comprehension. Accordingly, this level of vocabulary was associated with a passing grade of 56% on a test of reading comprehension for first-year university students taking a course in English for Academic Purposes (EAP). It was also found to cover 95% of the running words or tokens in short novels for younger readers (Laufer, 1989). In contrast, Hirsh and Nation (1992) demonstrated that knowing the 2,000 most frequent word families of English plus proper nouns allows low-intermediate second language readers to reach 95% of known word coverage. However, as mentioned earlier, this number is not enough to allow readers to determine the meaning of unknown words accurately. The studies by Laufer (1989, 1992) and Hirsh and Nation (1992) support Hu and Nation's (2000) findings that some learners will gain adequate comprehension of fiction texts with such a level of word knowledge coverage but that most need 98%. Nation (2006) also suggests that this threshold should in fact be raised to 98%, since 95% coverage is not sufficient to decipher and learn new vocabulary incidentally.

Given that there is very little research on incidental vocabulary acquisition through aural means, these 95% and 98% thresholds found in reading research have also been used in aural studies (for example, Adolphs & Schmitt, 2003, Brown, Waring & Donkaewbua, 2008, Nation, 2006). Based on these findings, I intend to use the more conservative aforementioned threshold of 98% to determine the availability of learnable

vocabulary in the various corpora. Following is a discussion of the language processing strategies used when inferring the meaning of unknown words, which support the use of the 98% known-word coverage threshold.

Language processing strategies

Deciphering the meaning of words from context in an ongoing manner may prove difficult even in one's native language. For example, Nagy, Anderson and Herman (1987) studying a group of elementary students reading in their first language (L1) – English – estimated the participants' success rate of incidental vocabulary learning at 5%. The task performed in the experiment seemed to be difficult even for older individuals. For example, Daneman and Green (1986) observed that even highly educated adults – university undergraduates – were often unable to accurately guess the meaning of non-words, i.e., words fabricated for the purpose of the experiment, encountered in a short paragraph when the context was not explicit. If such an endeavor as continuously guessing word meaning proves to be difficult for people of all ages in their native language, this exercise is expected to be, at the very least, just as difficult in a second language. Indeed, Paribakht and Wesche's (1999) study of vocabulary acquisition as a by-product of reading for comprehension found that intermediate-level university ESL students tended to ignore unknown words that were not essential for the general comprehension of the overall message. This leads us to acknowledge that, unless the words are repeated several times and the learner makes a conscious effort to pay attention to the unknown vocabulary and then infer a correct meaning, the chances that the words will be acquired (e.g., stored in memory for later recall) are small.

Another factor that could affect incidental L2 vocabulary acquisition is the transferability of processing ability from L1 to L2. Haastrup, Albrechtsen and Henriksen's (2004) preliminary results of their study involving Danish learners of English – grade 7 students and first-year university students – suggest that inferencing processes are different in L1 and L2. This becomes particularly important for adaptability, i.e., learners' ability to adapt their processing when faced with words that have linguistic cues to meaning as opposed to those that do not.

A similar study by Fraser (1999) investigated L2 learners enrolled in EAP courses with low and high levels of proficiency, and who were not taught strategies to find the meaning of unknown words. She found that the participants tended to ignore a greater number of those words compared to learners who were trained to use strategies such as consulting or inferencing through sense creation processing, i.e., creating meaning or defining words. She also found that they were just as successful using either one of these strategies separately. As expected, combining the two strategies appeared to enhance vocabulary acquisition.

Wesche and Paribakht (2010) note that, as previously mentioned, most unknown words go unnoticed by learners. They also point out that the occasions when L2 learners do notice unknown words are when those words are needed for text comprehension. In this case, inference is usually the preferred strategy. Their review of the literature also shows that interpretation is most successful when the clues needed are present in the word itself or in the adjacent text, and that the learners combine the information from the text with their general knowledge to construct meaning. Still, long-term retention of the words appears very limited even after appropriate inferencing. Their review of the

research on incidental acquisition of new lexical knowledge suggests that this serves as evidence for vocabulary acquisition being incremental and, consequently, it supports the claims that multiple encounters are needed to acquire word knowledge. They also highlight the importance of typological differences between languages for inferencing, and also mention that learners are more likely to transfer first language knowledge into their L2 if that language is close to the latter.

However, the influence of such transfer on inferencing remains largely unknown and this is what Wesche and Paribakht (2010) set out to investigate in a study involving the acquisition of English (L2) by L1 speakers of French and Persian. An introspective method – whereby research participants explained aloud what they were doing and what they were thinking – was used to determine and compare inferencing patterns between L1 and L2 of post-secondary students and to test learners' L2 word knowledge retention. Their findings are consistent with those of earlier studies. First, the results show that each group appears to have different patterns of knowledge source use in their L1. The three groups – the two English L2 groups and the English L1 group – use knowledge sources from all linguistic levels (i.e., word, sentence and discourse) and non-linguistic world knowledge sources, but to differing degrees. Sentence level inferencing is the most prominent strategy used in all three L1 languages but the other strategies appear to be used more in one L1 than others. For example, English L1 users tend to rely equally on word, discourse and world knowledge while French L1 users appear to depend more on word cues and Persian L1 readers, more on discourse cues. Regardless of their L1, the main strategy for learners working in their L2 appears to remain related to meaning knowledge sources, i.e., some type of inferencing strategy rather than, for example, word

retrieval or appeal for help (e.g., looking up a word in a dictionary) strategies. These knowledge sources are most often concentrated at the sentence level while the use of word cues remains prominent. The major difference observed between L1 and L2 is the knowledge source use: L1 readers rely more on *world knowledge*. Also, participants whose L1 was closer to the L2 reported using L1 features more often to infer meaning than did the participants whose L1 was more distant. Yet, participants from both groups demonstrated the integration of some L2 processing strategies more typical of L1 English than their own L1. For example, their use patterns of *word morphology* resemble more closely that of L1 English speakers than that of their respective L1 language. However, their overall use patterns of the various knowledge sources to decipher meaning of unknown words remain significantly different to that of L1 English speakers. It is not surprising then that the participants' success rate was significantly lower in their L2 compared to their L1, as both groups presented high failure rates: 78.0% for the Persian speakers and 48.4% for the French speakers. Even though some differences were observed between the Persian and the French groups (possibly attributable to the fact that the groups were assigned different sets of words to be tested), lexicalized words, i.e., words that have an equivalent translation from one language to the other were inferred correctly more often than non-lexicalized words (i.e., words that do not have an equivalent translation). These findings are not encouraging given that successful inferencing appears to be an important factor in initial L2 vocabulary learning. Significant but small vocabulary gains of target inferred words were observed as a result of their L2 inference tasks, but long-term retention rates remain unknown as the research did not test their effects.

In brief, the literature suggests that, in order to maximize vocabulary acquisition, learners need to be exposed to texts with a known-word coverage of 98% in which the unknown words are repeated several times, preferably in recurrent and varied contexts. Furthermore, learners need to be trained in noticing and making inferences about unknown words. These findings all relate to vocabulary encountered in written text and, accordingly, much less is known about acquiring words through listening. This will be the topic of the forthcoming discussions.

Conditions for incidental learning of new vocabulary through aural means

The scarce body of research on incidental vocabulary learning through aural means limits the conclusions we can draw on the subject. The forthcoming discussion will highlight the findings and shortcomings pertinent to the current research.

While listening to stories is one way to develop one's vocabulary (Elley, 1989, Nation, 2001, West, 1953), it does not necessarily prepare learners to understand and acquire vocabulary from speech. Up until now, research has allowed us to shed some light on some of the conditions necessary to gain vocabulary through aural means. An argument can be made for exposure to unknown words in a variety of contexts such as that offered by television program watching. Many other factors influence word learnability, which make pinpointing the exact number of repetition required for vocabulary acquisition difficult. The same is true for identifying the learner's vocabulary size necessary for this task. Certain advantages have been observed in favor of watching television compared to other means of aural exposure. A review of the relevant literature will support the choices made for the current research in these regards.

Listening to stories

Listening to stories has been shown to play an important role in developing L2 proficiency. West (1953) found that reading stories aloud to students helps improve their aural comprehension, while Elley (1989) and Nation's (2001) research yield evidence that substantial vocabulary gains from listening to stories are possible. Most studies on the subject involve children (e.g., Ewers & Browson, 1999; Penno, Wilkinson & Moore, 2002; Senechal, 1997) and so very few tested the effect of listening to stories on adults' vocabulary acquisition. One of the exceptions is the study by Brown, Waring and Donkaewbua (2008), who investigated vocabulary learning by a group of 35 Japanese university-students of English literature of pre-intermediate to intermediate level competence in English. They compared the retention rates through three different activities: reading, reading while listening and listening only. Although the results for reading and reading while listening were much higher than for listening alone (on the last post-test administered a month after the initial exposure the groups had a 15%, 16% and 2% retention rate respectively), the study suggests that listening without text support could have some effect on the acquisition of vocabulary. It also appears that learners need to encounter spoken words more frequently than the written ones in order to acquire them.

Factors influencing aural comprehension and vocabulary acquisition

Assuming that adequate text comprehension is also necessary for vocabulary acquisition through aural means, it is important to determine what factors influence aural comprehension and acquisition. Children's vocabulary acquisition of their first language is influenced by frequency of exposure interacting with different factors including

category, modality and more importantly, developmental stage (Goodman, Dale & Li, 2008). This interaction suggests that frequency of word occurrence is not the only factor that will affect L2 incidental vocabulary growth. Ellis (1995) investigated the effects of incidental vocabulary acquisition by providing modified oral input with a low-proficiency group of Japanese high school students enrolled in a listening and speaking English elective course. His results suggest that although comprehension, complete or partial, is necessary for vocabulary acquisition to occur, it does not guarantee that words will be acquired. It was also found that the range (i.e., the variety of contexts in which a word is used) and not the frequency of occurrence was the most important factor in the acquisition of new vocabulary. This suggests that encountering a given word in many different contexts helps strengthen the word-meaning relationship and thus its acquisition. Watching television programs may be a means to expose learners to a variety of contexts.

The number of studies on the frequency of word occurrence in oral text necessary for incidental vocabulary acquisition is very limited. Based on the current state of the field, however, it is reasonable to assume that the number of encounters needed to acquire a word is not inferior to the threshold established for word acquisition through reading, i.e., at least 6 encounters (Zahar, Cobb & Spada, 2001).

Word repetition in aural input

The need for higher vocabulary redundancy when acquiring words through speech discussed in Brown, Waring and Donkaewbua (2008) is not encouraging for those hoping to use listening as a vocabulary acquisition activity. The study found that to acquire vocabulary incidentally through listening alone, learners needed to encounter the

unknown words as many as 30 times. However, the authors argue that this particular group of learners displayed a lack of familiarity with spoken English and, because the phonological system of Japanese is considerably different from that of English, this could be one of the reasons why the results were so low for listening. The authors actually found that the learners did not remember any of the target words after three months, and only partial knowledge of a few words remained, i.e., when prompted, the participants were able to produce some of the correct answers, on average 8.2 of the 28 words. This is consistent with Fraser's (1999) findings that L2 learners acquire words faster if they have previously been exposed to them. Another reason for the disappointing results obtained in Brown et al. may be the learners' lack of skills in detecting word boundaries in connected speech. These constraints may have prevented the learners to focus on inferring the meaning of the unknown words the way learners did with the other techniques used in their experiment. Lastly, the learners were required to listen to a text with a 95% known-word coverage rate that is intended for vocabulary acquisition through reading and not listening. In a study of L2 listening comprehension Van Zeeland and Schmitt (2012) did corroborate the 95% mark in that 95% of known lexical coverage was adequate to understand the meaning of the different authentic, i.e. unscripted, spoken passages. However, no research to date has shown that 95% known word coverage is in fact sufficient for incidental vocabulary acquisition through listening. Indeed, a different rate may be necessary for learners to acquire words from listening (and not simply understand the message conveyed through oral means). Long-term retention results from Brown et al. also suggest that learners need to encounter a word significantly more than 20 times to acquire it through listening only. The results of the different types of test also suggest that

when participants receive some help to remember the vocabulary, more words appear to be learnt than when the participants are not prompted. This is consistent with previous research suggesting that word acquisition is incremental (e.g., Deighton, 1959, Elley, 1989, Nagy & al., 1985, Nation, 2001).

Previous research on incidental vocabulary acquisition through aural means has established more conservative thresholds to identify learnable words than the one suggested by Brown et al. (2008). Horst (2010), for instance, using the figure of 6 from Zahar, Cobb and Spada (2001) considered a word repeated 6 times or more in a teacher talk corpus of 120,553 words directed to adult learners to be learnable. In another study of teacher talk, now directed to younger learners, Horst (2009) established the word learnability level at 10 occurrences or more in a corpus of approximately 104,000 running words. Using a corpus much more limited in scope[1], Meara, Lightbown and Halter (1997) also argue that if a word is not encountered several times, it is not likely to be learnt. These authors, however, do not establish a minimum of encounters necessary for acquisition to occur, but they note that an average of only 2.75 unusual words were present per 500 words of running text; in addition, these words occurred only once or twice per 500-word segment. Zahar, Cobb and Spada (2001) arrived at similar conclusions and found that repetition needs for acquisition are related to learners' levels of proficiency. In other words, the previous knowledge and language ability of the learners seems to determine the number of word repetitions for a word to be acquired.

To summarize, a significant amount of research has been done with regards to word repetitions necessary for incidental vocabulary acquisition to occur. However, it is

---

[1] The corpus was only 9,500 words of teacher talk directed towards an audience similar to that of Horst (2009).

impossible to arrive at an exact number because of the different factors influencing word learnability, such as the nature of the word, the learner's L1 and level of proficiency, and the learning context. So far, most of the research on the subject involves incidental vocabulary acquisition through written texts. With regards to the number of encounters necessary for acquisition through listening specifically, it is fair to assume that learners need at the very least as many encounters in oral texts as with learnable vocabulary encountered through written texts. This is because listening increases the pace at which vocabulary exposure occurs, hence shortening the time for processing. Listening also usually prevents the learners from going back on the text to verify comprehension. These limitations suggest that learners need more known-word coverage in aural than with written texts to accurately guess the meaning of words in activities such as watching television. Hence, based on previous research (e.g., Meara et al. 1997; Brown, Waring & Donkaewbua, 2008; Horst, 2009 and 2010), the threshold for word learnability in the present research will be set at 10 encounters.

How many words do you need to already know?

In order to identify the learnable words available in each corpus, we must first determine the vocabulary already known by the learners. The results of previous studies provide some guidelines on how to proceed.

Much of the research on the subject of vocabulary size required for incidental vocabulary learning has been conducted with younger participants. Rice and Woodsmall (1988) argue that incidental vocabulary learning from television by children may be influenced in part by accumulated linguistic knowledge, prior viewing experiences, and the number of encounters with the target words. Oetting, Rice and Swank's (1995)

investigation of normally developed and language-impaired children also obtained results "[providing] some indication that the learners who knew the most words and had the greatest coverage learned the most words" (Webb & Rogers, 2009, p. 340).

Similar observations have also been found for adult participants. For instance, Bonk (2000) looked at the interaction between lexical knowledge and listening comprehension. The study involved 59 Japanese low-intermediate to advanced students of English. A recall test was used to evaluate listening comprehension and a dictation of the same text was used to gauge vocabulary knowledge.[2] The researcher's use of a written text read aloud (on a tape) is somewhat inauthentic but necessary to control the lexical difficulty of the discourse. The study showed that higher scores on the dictation correlated with better comprehension. Bonk was unable to determine a percentage threshold for achieving better comprehension. However, it was observed that some learners gained good comprehension of the text with only 70-79% knowledge of the tested words but that most required 80-100%. Interestingly, he found that these results are consistent across the different levels of learners' proficiency. These results suggest that learners need to know a substantial number of words from the text in order to understand spoken discourse. Based on this research, Bonk also suggested that learners be trained to recognize words in connected speech, since many participants had difficulty with this task. He stressed that lexical knowledge is only one of several elements playing a role in listening comprehension, thus leading us to hypothesize that the poor source of stimulus of the audiotape is in part responsible for the high percentage of lexical knowledge found

---

[2] The dictation was graded on the recognition of the lexemes and not on spelling or grammatical morphemes. In other words, knowing a word meant the participants knew its meaning. Because of its aural focus, grammar and spelling were not a concern for this particular research.

to be required for the understanding of spoken texts. In actual discourse, conversation is supported by the interlocutors' feedback just as television benefits from the visual support offered by the images. In this case, the coverage of known words might not, in fact, need to be as high, since other elements in the environment might contribute to comprehension and thus compensate for a lack of lexical knowledge.

Bonk did note that a substantial amount of vocabulary was necessary for learners to comprehend speech in their L2. However, the question of how much vocabulary is actually needed to understand spoken discourse remains to be clarified. Nearly 50 years after the study of Schonell, Meddleton and Shaw (1956), Adolphs and Schmitt (2003) revisited this question. The more advanced technological means at their disposition at the time allowed them to do a more thorough analysis than their predecessors. They used the Cambridge and Nottingham Corpus of Discourse in English (CANCODE), which contains approximately 5 million words, as well as the spoken component of the British National Corpus (BNC), composed of approximately 10 million words, both substantial in size compared to Schonell et al.'s modest corpus of 512,647 words (Adolphs & Schmitt, 2003). In addition to the sheer size of the corpora, the more varied sources of speech they provided made them more representative of general discourse than the corpus of their predecessors, i.e., Schonell et al.'s corpus. Adolphs and Schmitt tried to keep with Schonell et al.'s classification of the words as much as the technology that they were using allowed them in order to have a good basis for comparison. They observed that Schonell et al.'s conclusion that 2,000 word families providing around 99% coverage of spoken discourse was probably optimistic. In contrast, Adolphs and Schmitt found that 2,000 word families covered less than 95% of the CANCODE corpus and that an

additional 1,000 would raise the bar by only a little over 1%. The results were similar for the analysis of the BNC corpus. This suggests that one needs a much more elaborate lexicon to understand spoken discourse. As the authors point out, this is only speculation based on research involving written discourse.

Nation's (2006) study of the British National Corpus (BNC) supports these findings. He argues that knowledge of 98% of the vocabulary in a text is needed for unassisted comprehension. He found that this coverage translates into the knowledge of 8,000 to 9,000 word families for comprehension of written texts and 6,000 to 7,000 for spoken texts.[3] According to the author, this figure applies for unscripted spoken English as well as in scripted English such as is observed in *Shrek*, a children's movie. Given the knowledge of 7,000 word families, viewers of *Skrek* are exposed to one unknown word every 50 words, which offers great contextual support for comprehending new words and learning them. Thus it may not be more difficult to acquire vocabulary through oral input than it is with written text: Even though the discourse is generally presented at a faster pace than written text is, the processing demands on the learners might not be greater given that the vocabulary size necessary for comprehension is less than with written text.

A study by Van Zeeland and Schmitt (2012) suggests that 95% known word coverage is sufficient for listening comprehension. The participants (mostly university undergraduate students) listened twice to four anecdotes of unusual situations found on the Internet where the narrators used informal speech. Non-words were used to replace some of the words to control the known-word coverage and participants were made

---

[3] Nation used the lemmatized form of words to make the word count. It includes the base word (e.g., move) and its inflections (e.g., moves, moved, moving), as well as its basic derivative forms (e.g., mover).

aware of this fact so as to not hinder their concentration. Immediately after the first listening, the participants completed the multiple-choice test, which was comprised of 60 questions about the main ideas and details of the stories. They were allowed to review their answers after the second listening. The results show that adequate comprehension of the stories is reached with 95% known word coverage. However, no study has looked at the impact of the various levels of listening comprehension on incidental vocabulary acquisition. Given that the incidental vocabulary acquisition resulting from 95% known word coverage has yet to be determined, the current study will use a more conservative measure of 98% known word coverage to identify the learnable vocabulary in teacher speech and TV-talk.

This measure may also be more warranted given that the vocabulary demands from conversation and television are somewhat different. Viewers cannot clarify their understanding of the speech through interaction the same way interlocutors do. When watching television, the lexical context is very likely to influence the learners' comprehension of the message and, hence, the possibility of vocabulary acquisition. Webb and Rodgers (2009) found that the vocabulary demands of television programs are heavily dependant on its genres ranging from 2,000 to 9,000 word families to reach 98% coverage. This suggests that some genres are more appropriate for learners of lower proficiency levels while others may be more suitable for learners of higher proficiency. In turn, this will impact which words become learnable as the proficiency levels of the learners determine the 98% known-word threshold.

Oral input combined with visual support

An argument can be made for the use of television programs over other means of exposure to aural texts. Given that overall incidental vocabulary gains appear to be limited (e.g., Brown, Waring & Donkaewbua 2008; Elley, 1989; Nation, 2001), teachers will want to maximize the learning opportunities of their students.

Some studies have found that learners acquired more words through story telling when the words were associated with an image (e.g., Elley, 1989, Jones & Plass, 2002, Mueller, 1980). For instance, Mueller (1980) observed that visual supports enhanced comprehension recall in adult beginner learners of German, especially if these visual aids were presented before hearing the recording. Similar results were obtained in a study of 8-year olds English native speakers by Elley (1989), where it was found that learners acquired more words through story telling when the words were associated with an image than when an explanation of the word was given. However, when replicating his study with more advanced learners, Mueller failed to obtain similar results, which suggested that visual supports were most beneficial for beginners. In other words, the level of proficiency appears to determine the conditions under which learners best comprehend speech, i.e., beginners appear to benefit more from visual support. In the context of television, where speech is always associated with images, one can then hypothesize that proficient learners trying to acquire words incidentally may be able to compensate for their lack of vocabulary by the interpretation of images viewed.

There are also studies that have found that offering a combination of visual and written supports with listening could increase vocabulary acquisition (e.g., Brown, Waring & Donkaewbua, 2008, Jones & Plass, 2002). Jones and Plass' (2002) research

involving English college students enrolled in a beginner French course showed that learners acquired more vocabulary through listening and remembered the stories better when combining written and pictorial annotations than when choosing only one or none of these techniques. They also found that visual supports yielded more important vocabulary gains than written annotations. The authors suggest that the multiple retrieval cues that the combination of these techniques offers allow the learners to acquire more words and thus to commit more vocabulary to long-term memory. Similarly, Brown, Waring and Donkaewbua (2008) have shown that reading-while-listening is the most effective method for incidental word acquisition. Hence, the fact that oral input combined with visual and written supports helps increase vocabulary acquisition suggests that extensive exposure to television programs could be beneficial to second language learners as the medium offers both visual support and often captions for hearing impaired viewers.

The case of teacher talk

Students are exposed to a substantial amount of oral input in the classroom and, consequently, teacher talk remains an important part of L2 instruction. Research shows that this type of input does provide learners with the opportunity to acquire vocabulary (Horst 2009, 2010; Meara, Lightbown & Halter, 1997; Vidal 2003; Wode, 1999).

Wode (1999), for example, compared the vocabulary acquisition of English as a Foreign Language (EFL) immersion students to regular EFL students at a school in Germany. He observed that the students in the immersion group greatly outperformed the students in the other group on a vocabulary test, most possibly due to the quantity, and not the quality, of the input provided in the immersion environment. Interestingly, Wode

noted that the majority of the words acquired by the students came from the teacher's speech as most of the terms learnt were not in the textbooks used in class.

Most research on the effects of teacher-based input is done in non-immersion settings, which could explain why they reveal much more limited gains. Vidal (2003) investigated the vocabulary acquisition of first-year university students of various proficiency levels (participants had TOEFL scores ranging from 387 to 661 with a group average of 507) through academic listening in an EFL course. The author found that learners' proficiency levels affected the rates of vocabulary acquisition. The reported gains appeared to be more significant among learners who demonstrated high language proficiency. She observed that more proficient learners were able to comprehend more conceptual information and thus retained more technical terms compared to less competent learners who had to dedicate more time and effort to understanding the main concepts, focusing less on unknown words. She also observed that although frequency of occurrence influenced vocabulary retention, other factors such as predictability from word forms and parts, word type and type of input elaboration by the teacher led to better learning.

Based on these studies, it becomes relevant to explore the vocabulary available in teachers' speech. In other words, what exactly has teacher talk to offer in terms of vocabulary? Which and how many words are available for uptake? Meara, Lightbown and Halter (1997) investigated the lexical richness provided by teacher talk and observed that when this lexical richness was measured in terms of low-frequency words, the environment was, in fact, quite poor. However, a closer look at what learners actually knew revealed that most probably they did not know a lot of the high-frequency words

and that, in fact, the opportunity for vocabulary acquisition provided by teacher talk was quite high. The authors concluded that if the learners fail to acquire vocabulary, the reason might not be the lack of availability of the learnable words but rather that these people fail to identify, understand and consequently acquire the terms.

About a decade later, Horst (2009) revisited this claim and found that lexical richness was still present in the classroom environment. This study also found that certain important words associated with particular themes were never encountered in teacher talk. For instance, many words linked to controversial topics such as armed conflicts were avoided in teacher speech. In addition, many important terms related to business and commerce, the physical world, history and social sciences were also omitted. By studying a teacher talk corpus of 121,000 words addressed to advanced adult learners of ESL, Horst (2010) found that with the knowledge of 4,000 frequent words, learners would be able to understand 98% of the vocabulary present in the teachers' speech. She also noted that very few unfamiliar words were sufficiently repeated for long-term retention by the students. Where will learners have the opportunity to acquire these unfamiliar words? The answer might be by being exposed to television programs.

The case of TV

In the previous section, it was observed that the classroom's lexical richness found in teacher talk is deficient in some areas such as controversial topics, business and commerce, social sciences and the physical world. It is thus possible that L2 learners could benefit from multiple and varied exposure to the spoken L2 language which could be provided by television programs. Television watching is in fact one of the most popular activities among the general public, so this experience might be of interest to

those wishing to complement classroom instruction with extra-curricular activities that may contribute to learning. This will be the topic of the discussion that follows.

TV-based listening vs audio-only listening

Researchers found that L2 learners retain more information and learn more vocabulary by combining aural input with some form of visual aid than by being solely exposed to audio messages. For instance, Tyson (1990) showed that intermediate learners of ESL are able to recall more information when watching a video than when only listening to the audio version of the same video. Similarly, Duquette and Painchaud (1996) studied the acquisition of French L2 vocabulary through exposure to oral texts and compared the results of using visual aids versus relying solely on the audio medium. The 119 participants who demonstrated a high-elementary proficiency level were divided into four groups: the experimental video group, the experimental audio group, the control audio group, and the control group without treatment. They watched or listened (depending on the experimental group) to an 8-minute clip from the National Film Board of Canada entitled *Le permis de conduire* ('*The Driver's License'*) twice and were tested on 40 target words. These items were presented in a sentence that was repeated twice and the participants had to give a translation of the words. A first post-test was administered immediately after treatment, followed by a second test, 10 days later. The authors found that more unfamiliar words were learned through the video treatment than the audio treatment. They concluded that the main factors affecting vocabulary retention are redundancy and variability of contexts. In other words, the number of times a word occurs in similar linguistic and visual contexts determines to a large extent the likelihood of its acquisition. These results suggest that TV-based homework could be more

beneficial than other audio-related activities such as merely listening to an aural-only input.

In a related study, Meara (1993), contracted by the British Broadcasting Corporation (BBC), examined the input, rather than actual vocabulary learning, of a number of broadcasts developed as part of the BBC Core Curriculum. He evaluated the levels of difficulty, in terms of the lexical load. After a first analysis, Meara was unable to show that the levels of difficulty provided in the different programs were compatible with the proficiency levels of the targeted learners. However, a comparative analysis allowed him to demonstrate that spoken and written texts are substantially different from one another, as are pedagogical and non-pedagogical materials such as those found in TV broadcasts. He also showed that different types of texts have different levels of difficulty. For example, he mentioned that Chris de Burgh's songs are much easier to understand than the pedagogical broadcasts produced by the BBC, which, on the other hand, are far from providing the lexical richness offered by the comic-strip books for young readers, *Tintin* (Meara, 1993, p. 36). His findings suggest that learners could benefit from combining classroom learning and exposure to television programs because the latter may provide a richer, or at the very least a complementary, lexical environment for learning to take place. The goals of this study are to determine the potential lexical benefits of listening to teacher talk, TV-talk and of combining these two activities.

Learning vocabulary from TV

It has been shown that repeated exposure to television programs yields some vocabulary acquisition in L2 learners. Results of previous research also suggest that particular programs and certain viewing conditions, such as the use of closed captions,

produce better results. As such, Uchikoshi (2006) looked at the effects of educational television viewing by Spanish-English bilingual kindergarten children on their receptive and expressive vocabulary acquisition. Students watched 30-minute episodes three times a week in class. The vocabulary was not reinforced by any activities after the viewings. Even after a whole year of this treatment, there were no vocabulary gains from the viewing of the programs. However, some gains were observed in those students who combined classroom viewings with home screening of the same programs. Results also differed depending on the show that students were watching, which suggests that not only do learners need repeated exposures to acquire vocabulary, but also that certain shows are more beneficial than others. Determining what those programs are could potentially increase learners' incidental vocabulary acquisition. Along these lines, the present study aims at uncovering which television genres (drama, situational comedy or science fiction) are most useful for vocabulary acquisition.

A similar study was conducted by Vanderplank (1993), who promoted the use of television programs with closed captions (same language subtitles) in the language classroom, a scenario in which content is conveyed by words, not images. The latter only support the message but fail to communicate the ideas fully. Vanderplank argues that watching television frees the teacher to some extent, offers the students a variety of choices and, more importantly, assigns responsibility to the learners where they take control over their own learning. He acknowledges, however, the existence of problems in exploiting scripted verbal messages in television programs. He identified the fast pace at which the language is provided, the quantity and nature of the speech, i.e., formal or colloquial, and the fact that language is culturally bound. To circumvent these problems,

Vanderplank suggested using closed captioning because it increases redundancy and provides a synopsis of the dynamic of the speech thus reducing the cognitive demands made on the learner. He also pointed out that certain programs work better than others for language learning. However, his position relates to the students' interest in the subject matters rather than to the vocabulary difficulties that the programs present. The incidental vocabulary opportunities offered by the various genres of television programs is still unknown. The current research aims to clarify this point.

Many studies have investigated the effect of television watching combined with captions – same language transcriptions – and subtitles – translations (e.g., Koolstra & Beentjes, 1999; Koskinen, Knable, Markham, Jensema & Kane, 1996; Ydewalle & Van de Poel, 1999). These studies have found that even without the use of these tools, vocabulary acquisition from watching television program was possible. Based on the results obtained in their study, D'Ydewalle and Van de Poel (1999) suggest further that language acquisition from movies, subtitled or not, is present only at the vocabulary level, not at the syntactic or morphological grammar levels. Hence, the reality remains the same: a variety of tools can be used to enhance incidental vocabulary acquisition while watching television, but multiple and varied exposures to unknown words need to be part of the learners' input for acquisition to take place. This will be the topic of the following discussion.

Input availability

Previous research suggest that "the combination of visual imagery and aural input may make it easier to learn words incidentally through watching television programs than through reading" (Webb & Rogers, 2009, p. 341) (see also Elley, 1989; Horst, Cobb &

Meara, 1998 for a similar view). Following these observations, researchers set out to determine the vocabulary demands and availability of a variety of television environments. This is exactly what Webb and Rogers (2009) examined in their study. Looking at the vocabulary demands of watching British and American television news and programs, the three criteria for program selection was availability of transcripts, genre, and popularity. Although the corpus was somewhat opportunistic and the sub-corpora unevenly compiled, Webb and Rogers obtained interesting results[4]. First, they looked at the number of known words necessary to understand the vocabulary in TV programs. Then, they calculated how word coverage differed between programs and different genres for British and American television. Finally, they examined the level of exposure to low-frequency words by the people watching those programs. The authors found that the lexical demands of the various genres in both corpora were similar, requiring 2,000 to 4,000 known words families plus proper nouns to reach 95% coverage and 5,000 to 9,000 plus proper nouns for 98% coverage. In fact, British shows appear to require fewer words and make similar vocabulary demands across programs – 6,000 to 7,000 word families. Comparatively, American shows seem to offer a greater disparity of lexical demands ranging from 6,000 to 9,000 word families (this excludes children's programs, which are produced with the intent of being easier to understand). They also

---

[4] Some sub-corpora were bigger than others. In fact, the British corpus was divided in two sections, the news and TV programs, because the authors could not find enough programs to justify using the six sub-categories present in the American corpus. There were 193,460 running words for the American corpus and only 70,924 for the British one. For the American corpus, dramas consisted of 53,922 words, the situation comedies, 31,201 words, and the children's programs, only 13,090. Such a big discrepancy in the total word count could well have influenced the data with regards to the low frequency words. Indeed, the variety of low-frequency words encountered is likely to be limited by the small size of the British corpus, so a direct comparison between the British and the American corpus at this level might not be accurate.

found that the lexical demands varied greatly from episode to episode, even within a single genre. This suggests that teachers should be careful about choosing single episodes for their learners, as random selection is not likely to yield the level of difficulty representative of the genre of a particular TV show.

In the same study, more than half (55%) of the low-frequency words were encountered only once, while 12% were encountered at least 5 times. This could lead to significant vocabulary acquisition over a year given that the learners watched an hour of TV a day. Webb and Rodgers also pointed out that if knowledge of the 3,000 most frequent words is sufficient for comprehension, the learners at this level are more likely to learn the words at the next 1,000 word levels. This is consistent with Krashen's $i + 1$ theory (Krashen, 1985), which suggests that learners who know the 3,000 most frequent word families stand to gain the most vocabulary from watching television. Webb and Rodgers observed that after the 3,000 word family mark, the greatest concentration (30%) of low frequency words repeated 5 times or more is at the 4,000-word level. The number of repetition for low frequency words then decreases as frequency decreases. In other words, the cognitive load appears to be too great for beginner learners, and advanced learners will need to watch a considerable number of hours of TV in order to gain new vocabulary as repetition of unknown vocabulary items becomes more and more scarce as learners acquire more lexical items.

Choice of television programs

Of what is currently available on television for adult entertainment, I propose that sitcoms, dramas, and sci-fi programs are the most popular given that they occupy the most popular time-slots (Gitlin, 2000). These genres have been selected for analysis in

this study because they provide the greatest coverage of television programs potentially watched by English L2 learners.

Concluding remarks and research questions

This review of the literature began by highlighting studies of incidental vocabulary acquisition that set out to provide evidence for Krashen's theory of learning new linguistic features through exposure to comprehensible input. These studies showed that incidental vocabulary acquisition is in fact possible assuming that learners will encounter the words enough times and that their L2 proficiency levels are such that the density of known words remains over 95% or, ideally, over 98%. If the density of unknown words is too great, the learners will not be able to use their processing strategies effectively to infer and acquire the meaning of the unknown terms.

It is important to note that these early investigations focused largely on learning vocabulary through reading. There are significantly fewer studies on the acquisition of new vocabulary through spoken input and none that conclusively examines the vocabulary learning potential of exposure to large amounts of L2 television viewing. The studies of incidental vocabulary acquisition through attending to spoken input – with and without images – that do exist suggest that this is a potentially useful resource for incidental vocabulary acquisition in itself, but also as a supplement to classroom oral input.

Recently, corpus based studies have attempted to delineate the potential for vocabulary learning exposure to large amounts of oral input. Speech that has been explored in this way includes story-telling, audio-listening, television viewing and teacher talk. Webb and Rogers' (2009) study looked at the learnable vocabulary available

on television. However, their corpus was unevenly compiled and, consequently, the results regarding low-frequency vocabulary items are likely to have been affected by this limitation. In this study, I address the comparability issue and account for the sizes of the corpora. In addition, I further the investigation of Webb and Roger with regards to the difference in potential benefit offered by the various genres of programs. I know of no study examining the potential lexical uptake value of combining teacher talk and television programs.

As demonstrated above, there is a limited body of research about incidental vocabulary acquisition through aural input, especially with regards to the potential value of using TV-based homework. To address this gap in the literature, I propose to answer the following questions in the present study:

a. How does the vocabulary found in teacher talk compare to vocabulary found in TV-talk in terms of words used and word frequency?

b. In the case of TV-talk, how does the learnable vocabulary compare among the three genres considered in this study: sitcoms, dramas, and sci-fi programs?

c. What/how many words in teacher talk are more likely to become learnable when combined with TV-based homework?

Chapter 3 – Methodology

This study aims to determine the vocabulary acquisition potential of the various genres of programs available during prime-time TV, as well as the added potential of combining these genres with teacher talk. In order to answer the questions put forward in the previous chapter, this section describes the methods utilized in this study. It starts by outlining the criteria and methodologies used to compile the teacher talk and TV-talk corpora. Finally, it provides the steps followed through the analysis to arrive at the results described in the next chapter.

The teacher talk corpus

The teacher talk corpus was collected in a class of adult high-intermediate to advanced ESL students. The course was given through a community centre and all learners were recent immigrants. The teacher was a native English speaker with training in applied linguistics and 7 years of teaching experience teaching abroad. She was unaware of the goal of the study. The instruction was focused on communicative activities concentrating on speaking and listening skills. Since there were no vocabulary tests, it is fair to assume that most words were learnt incidentally although students might have taken notes and studied certain words on their own. The classes were held in a research laboratory facility in Montreal twice a week for 9 weeks and each class lasted approximately 2 hours. The teacher wore a Lavaliere microphone to ensure sound quality. The corpus and related transcripts were compiled for a research project from Concordia

University (ALERT project[5]; Collins, Trofimovich, Horst, White & Cardoso, 2009).
Those transcripts represent 32 hours of class time in total. For the purpose of this
research, all student-produced speech was excluded. The teacher talk also contained some
scripted speech: a song, some lines read aloud from a textbook and a dictation. In the
present study, those were taken out to keep only the spontaneous speech produced by the
teacher.

A spellchecker was used to identify unconventional spelling and transcription of
connected speech (e.g., "hafta") and changes were made to standardize the text. For
example, speech-fillers such as *hummm* or *huh* were regularized to *hum* and *uh*.
Connected speech such as *hafta* were also transcribed as *have to*. This was done so that
the frequency software would not classify highly frequent words such as *have to* as rare
words because of their unusual spelling. It was assumed that high-intermediate learners
can decode connected speech and recognize that *hafta* is in fact the combination of two
words: *have* and *to*.

Each class was saved in a different file and each of these files contain between
1,082 and 4,886 token words of teacher talk for a total of 111,812 words for the entire
corpus.

The TV corpora

The TV corpora are composed of three different sub-corpora: the sitcom corpus,
the drama corpus and the sci-fi corpus. The sitcom and drama sub-corpora have been
compiled by a group of students as part of an Applied Corpus Linguistics MA course at

---

[5] The *Alert project – The sounds, the forms, the meanings: An integrated perspective on
second language teaching and learning*: was funded by the *Social Sciences and
Humanities Research Council of Canada* (*SSHRC*).

Concordia University during the winter 2009 term, while the sci-fi sub-corpus has been compiled specifically for this research project by the researcher.

The data collection for the sitcom and drama sub-corpora depended on the availability of the transcripts on the Internet. The episodes were transcribed and uploaded on the web by fans of the shows. Each sub-corpus is composed of 5 different programs, each TV program file is around 50,000 words, which amounts to around 10 hours of television viewing. The length of drama programs is an hour and the sitcoms are 30 minutes long; consequently, more episodes were required for the sitcom sub-corpus to reach equivalence between the two genres. The sitcom sub-corpus includes transcripts from: Frasier, How I Met your Mother, The Office, Seinfeld and Two and a Half Men. The files range from 38,129 to 69,784 words (tokens) for a total of 259,203 words. The drama sub-corpus is composed of transcripts from: Alias, Desperate Housewives, Grey's Anatomy, Lost and Prison Break. The files range from 50,909 to 59,450 words for a total of 264,898 words for the entire drama sub-corpus. The transcripts were stripped of all stage directions manually and another student looked over the work to ensure the quality of the sub-corpus.

The sci-fi sub-corpus includes the following shows: Battlestar Galactica, Doctor Who, Fringe, Stargate SG-1 and Star Trek – The Next Generation. These programs, as the ones used in the sitcom and drama sub-corpora, were airing for the first time or had known a great success and reruns were still available on television at the time of data collection. The selection of programs was made using the *Top 100 Sci-Fi TV Shows List* (retrieved July 28, 2010 at http://home.austarnet.com.au/petersykes/topscifi/lists_ tv.html). In addition to still being aired on TV at the time of the selection, the other

criterion for program selection was the availability of its transcript on the Internet.[6] This

was done to make sure that the data collected would be representative of what is available

on TV at the time the research was conducted. To ensure their quality, two episodes of

each program have randomly been selected and a viewing of these shows has confirmed

the accuracy of the transcripts. As was the case with the sitcom and sci-fi program, the

transcripts were stripped of all stage directions. The durations of the episodes are between

30 minutes and 1 hour, which amount to approximately 10 hours of viewing for each

program, for a total of about 50 viewing hours for the entire sci-fi sub-corpus. The files

for each programs range from 48,574 to 51,620 words and the total number of words for

the sci-fi sub-corpus is 250,067.

Analysis

The corpora were analyzed using Nation's (2006) BNC-based frequency lists and

corpus tools developed by Cobb (2010) and available on lextutor.ca. To determine the

levels of coverage offered by each of the 20 levels of Nation's frequency lists in the

teacher talk corpus and the TV corpora, each corpus was analyzed through the

Vocabprofile BNC-20 program. Proper nouns (show characters such as *Charlie* and *Alan*)

were reclassified from less-frequent words to the first 1,000 word-family list because

learners were expected to know them. In addition, the spelling of speech fillers and

interjections such as "hum", "huh" or "oh" were standardized (e.g., "hhuh", "huhhhh",

"heh" will be changed to "uh") so that they could be recognized by the program and

---

[6] Transcripts for Battlestar Galactica, Doctor Who, Fringe (episodes 1 to 9 of season 1)
and Stargate SG-1 have been obtained through http://www.twiztv.com/scripts, while
episodes 10 to 20 (season 1) of Fringe were collected from http://fringepedia.net/wiki and
Star Trek – The New Generation transcripts are from http://www.chakoteya.net/
NextGen/episodes.htm.

classified as first-1,000 level words and not as off-list words – words that were not in the lists of 20,000 frequent words of English. Most connected speech forms such as "can't, won't, isn't, shouldn't" were automatically broken down into their different components (e.g., "won't" became "will not") and classified as first-1,000 level words. Words such as "cos" and "gonna" were kept intact and only reclassified as first-1,000 level words because they are frequently used words characteristic of speech. Words like "d'you" saw the "d" deleted and the "you" classified as first-1,000 level words by the program and so they had to be manually changed back to their conventional spelling in the transcript so that these words were not lost from the frequency computation. The same applied to "ain't" because the vocabprofiler broke it down into "ai", an off-list word, and "not" a first 1,000 level word. A number of inconsistencies in the interjections and proper nouns remained after having gone through the program and so were removed manually for the analysis.

Educational institutions each have their own way of classifying students' levels of proficiency, and so do researchers. As there is no consensus as to what a beginner, intermediate and advanced learner is in terms of vocabulary, herein Nation's (2001) findings were used to categorize levels of proficiency. According to his research, native speakers know around 20,000 word families. In addition, the first 2,000 word families cover almost 80% of the running words in texts; beginners should know this high frequency vocabulary first. 3,000 word families are also shown to be necessary to understanding authentic texts without any help from dictionaries or other resources. Accordingly, knowing 3,000 to 4,000 is an acceptable goal for intermediate learners. Consequently, advanced learners would be those who know 5,000 words or more. Based

on this classification, it was then possible to determine the level of proficiency necessary for an L2 learner to comprehend the text, and how much vocabulary was available for uptake. In other words, this classification allowed us to determine how many words learners need to know in order to understand a particular genre of programs, i.e., to reach 98% of known word coverage, the threshold for unassisted oral text comprehension (Nation, 2006; Webb & Rogers, 2009; Horst, 2010). Using the text-based Range software (also available at lextutor.ca), I was able to establish how many unknown words are repeated enough times to be learnt – i.e. 10 times or more, as discussed in Chapter 2 – in the corpus and/or in combination with teacher talk.[7] This software allowed us to determine which and how many words can be acquired from listening to teacher talk and/or the various TV corpora. Following this procedure, the data for teacher talk was compared with those of the TV corpora to see where there were overlaps and to determine what vocabulary items are unique to a specific genre. By combining the teacher talk corpus with each TV corpus, it was possible to determine how much more vocabulary can be acquired by combining the two activities. This way, I was able to establish which words met the required threshold for learning to take place (in isolation or in combination), and what type of TV programs had more to offer as TV-based homework for the improvement of vocabulary acquisition. This was made possible by using word count and percentages.

---

[7] As no consensus is formed yet on the number of repetitions necessary for incidental vocabulary acquisition to occur, a minimum of 10 occurrences has been determined based on the literature available on the subject up to this point (Meara et al., 1997; Brown, Waring & Donkaewbua, 2008; Horst, 2009 and 2010) – see previous discussion in Chapter 2.

Chapter 4 – Results

This chapter presents the findings to the three research questions introduced in Chapter 3 and, as such, it is organized based on these questions. First, the learnable vocabulary of teacher talk is compared to that of TV-talk in terms of words used and word frequency. Then, closer attention is paid to individual genres of TV-talk (sitcoms, dramas and sci-fi programs) to identify the vocabulary learning opportunities of each. The chapter ends with a report of the added incidental vocabulary learning benefits when one combines exposure to both teacher talk and TV-talk.

How does the vocabulary found in teacher talk compare to the vocabulary found in TV-talk in terms of words used and word frequency?

The learnable vocabulary – i.e. words found after the 98% known-word coverage threshold and repeated 10 times or more in a corpus – found in teacher talk was compared to that of TV-talk. The purpose was twofold. First, the goal was to determine whether the words used in one corpus overlapped with the learnable vocabulary used in the other, or if the two environments (teacher talk and TV-talk) complement each other. The second objective was to establish whether one source of input is easier to understand (i.e. requires a smaller known vocabulary) and more beneficial to the learners (i.e. provides a greater amount of learnable words). This was determined based on word frequency – using the BNC-20 classification from the Vocabprofiler on lextutor.ca. My results show that teacher talk requires the learners to know fewer words, than TV-talk does to reach the 98% known-word threshold necessary to decipher new words, i.e. the 4,000 most frequent words for TV-talk compared to 7,000 to 8,000 high frequency words for TV-

talk, as shown in Tables 1 through 4 for the complete BNC word-levels analysis of vocabulary in teacher talk, sitcoms, dramas and sci-fi programs). Also, some overlap can be observed in the BNC word-levels in Table 5 at the BNC levels 8 through 11, which means that learnable words are present at the same BNC word-levels for both types of talk. However, Table 6 shows that the specific teacher talk and TV-talk learnable word items do not overlap and are thus complementary to one another. For example, learnable words are present in both TV-talk and teacher talk at the BNC-11,000 word-level however, the learnable items they provide are different. TV-talk provides the learnable words: *closet*, *freckle*, *fugitive*, *genocide*, *hangar*, *lore* and *teal*, and teacher talk offers the learnable words: *elf*, *scowl*, *slink*.

Table 1 – Lexical Frequency Profile of the Teacher Talk Corpus

Frequency level	Family	Type	Token	Coverage (%)	Cumulative (%)
K1 words	980	1,971	107,565	93.04	93.04
K2 words	571	865	3,444	2.98	96.02
K3 words	303	397	1443	1.25	97.27
K4 words	182	219	855	0.74	**98.01**
K5 words	122	156	545	0.47	98.48
K6 words	76	87	293	0.25	98.73
K7 words	63	76	296	0.26	98.99
K8 words	43	52	177	0.15	99.14
K9 words	34	40	122	0.11	99.25
K10 words	25	30	129	0.11	99.36
K11 words	23	27	88	0.08	99.44
K12 words	13	13	43	0.04	99.48
K13 words	10	11	44	0.04	99.52
K14 words	5	6	24	0.02	99.54
K15 words	11	11	32	0.03	99.57
K16 words	9	11	23	0.02	99.59
K17 words	2	2	6	0.01	99.60
K18 words	2	2	5	0.00	99.60
K19 words	5	7	20	0.02	99.62
K20 words	2	2	13	0.01	99.63
Off-list words	?	226	441	0.38	100.00
Total	2481+?[a]	4314	115,608[b]	100	100

[a] The Vocabprofiler program is unable to group words of the same families that are not on the frequency lists. This is why the question mark appears for the off-list words.
[b] The number of token words is different from the one previously mentioned because of how the Vocabprofiler counts contracted forms such as *didn't*. In these cases, the program counts them as two words: *did* and *not*.

Table 2 – Lexical Frequency Profile of the Sitcom Corpus

Frequency level	Family	Type	Token	Coverage (%)	Cumulative (%)
K1 words	2,164	4,073	247,667	90.14	90.14
K2 words	911	1,966	10,241	3.73	93.87
K3 words	766	1,284	4,430	1.61	95.48
K4 words	611	897	2,733	0.99	96.47
K5 words	470	615	2,112	0.77	97.24
K6 words	360	448	1,217	0.44	97.68
K7 words	271	314	802	0.29	97.97
K8 words	237	270	508	0.18	**98.15**
K9 words	220	246	429	0.16	98.31
K10 words	164	174	416	0.15	98.46
K11 words	131	150	252	0.09	98.55
K12 words	99	104	197	0.07	98.62
K13 words	102	105	175	0.06	98.68
K14 words	59	65	99	0.04	98.72
K15 words	89	103	175	0.06	98.78
K16 words	74	81	159	0.06	98.84
K17 words	64	69	126	0.05	98.89
K18 words	49	55	79	0.03	98.92
K19 words	50	54	280	0.10	99.02
K20 words	22	25	62	0.02	99.04
Off-list words	?	1,436	2,594	0.94	100.00
Total	6,913+?[a]	12,294	274,753[b]	100	100

[a] The Vocabprofiler program is unable to group words of the same families that are not on the frequency lists. This is why the question mark appears for the off-list words.

[b] The number of token words is different from the one previously mentioned because of how the Vocabprofiler counts contracted forms such as *didn't*. In these cases, the program counts them as two words: *did* and *not*.

Table 3 – Lexical Frequency Profile of the Drama Corpus

Frequency level	Family	Type	Token	Coverage (%)	Cumulative (%)
K1 words	2,051	3,919	254,187	90.85	90.85
K2 words	897	1,926	10,692	3.82	94.67
K3 words	751	1,228	4,412	1.58	96.25
K4 words	580	820	2,462	0.88	97.13
K5 words	440	584	1,456	0.52	97.65
K6 words	305	382	845	0.30	97.95
K7 words	227	259	618	0.22	**98.17**
K8 words	209	243	497	0.18	98.35
K9 words	180	207	377	0.13	98.48
K10 words	141	156	317	0.11	98.59
K11 words	104	117	326	0.12	98.71
K12 words	86	98	237	0.08	98.79
K13 words	83	97	208	0.07	98.86
K14 words	59	63	110	0.04	98.90
K15 words	67	73	143	0.05	98.95
K16 words	59	61	115	0.04	98.99
K17 words	44	51	125	0.04	99.03
K18 words	31	31	39	0.01	99.04
K19 words	40	45	65	0.02	99.06
K20 words	24	25	55	0.02	99.08
Off-list words	?	1,329	2,511	0.90	100.00
Total	6,378+?[a]	12,794	279,797[b]	100	100

[a] The Vocabprofiler program is unable to group words of the same families that are not on the frequency lists. This is why the question mark appears for the off-list words.
[b] The number of token words is different from the one previously mentioned because of how the Vocabprofiler counts contracted forms such as *didn't*. In these cases, the program counts them as two words: *did* and *not*.

Table 4 – Lexical Frequency Profile of the Sci-fi Programs Corpus

Frequency level	Family	Type	Token	Coverage (%)	Cumulative (%)
K1 words	1,908	3,927	231,213	87.58	87.58
K2 words	918	2,160	13,784	5.22	92.80
K3 words	746	1,331	5,411	2.05	94.85
K4 words	593	940	3,629	1.37	96.22
K5 words	467	664	2,121	0.80	97.02
K6 words	345	460	1,409	0.53	97.55
K7 words	276	323	809	0.31	97.86
K8 words	262	322	848	0.32	**98.18**
K9 words	218	252	501	0.19	98.37
K10 words	186	214	596	0.23	98.60
K11 words	151	163	354	0.13	98.73
K12 words	119	135	358	0.14	98.87
K13 words	105	121	226	0.09	98.96
K14 words	87	97	238	0.09	99.05
K15 words	51	58	112	0.04	99.09
K16 words	31	32	53	0.02	99.11
K17 words	49	54	86	0.03	99.14
K18 words	38	38	67	0.03	99.17
K19 words	37	40	83	0.03	99.20
K20 words	21	22	36	0.01	99.21
Off-list words	?	1,018	2,070	0.78	100.00
Total	6,608+?[a]	13,306	264,004[b]	100	100

[a] The Vocabprofiler program is unable to group words of the same families that are not on the frequency lists. This is why the question mark appears for the off-list words.
[b] The number of token words is different from the one previously mentioned because of how the Vocabprofiler counts contracted forms such as *didn't*. In these cases, the program counts them as two words: *did* and *not*.

This is consistent with the fact that comprehension of teacher talk is less demanding than TV-talk. In fact, as previously mentioned, the comprehension threshold necessary to acquire new words, i.e. 98% known-word coverage, is reached at the BNC-4,000 word-level for teacher talk and at the BNC-7,000 to 8,000 word-levels (depending on the genre) for TV-talk. Thus, teacher talk seems to be more advantageous for intermediate learners while TV-talk proves to be more beneficial for advanced learners as per the data used in this research, based on the learners' proficiency levels established in Chapter 2 (i.e., beginners are expected to know 2,000 words, while intermediate and advanced learners should know 3,000-4,000 and 5,000+ respectively).

Table 5 – Number of Learnable Words per BNC-1,000 Word-Level

for Sitcom, Drama, Sci-Fi, TV-Talk and Teacher Talk

BNC	Sitcom	Drama	Sci-Fi	TV-Talk*	Teacher Talk
BNC-5					8
BNC-6					8
BNC-7					9
BNC-8		6		6	3
BNC-9	3		9	12	3
BNC-10	2	2	12	16	5
BNC-11	1	2	4	7	3
BNC-12	2	4	9	15	
BNC-13	2	2	4	8	
BNC-14	1	2	5	7	
BNC-15	1	1	1	3	
BNC-16	2	2		4	
BNC-17	1	1		2	
BNC-18			1	1	
BNC-19			1	1	
BNC-20	2	1		3	1
OFF-LIST	12	6	22	37	2
Total number of Learnable Words	29	29	68	122	42

*Certain words are repeated in different genres and are only counted once in TV-Talk.

Table 6 – Learnable Words in Sitcoms, Dramas, Sci-Fi Programs and Teacher Talk for

BNC 1,000 Word-Levels 8 through 11

BNC 1,000 Word-Levels	Sitcoms	Dramas	Sci-Fi Programs	Teacher Talk
8		Attorney Dynamite Raft Ranch Spank Surveillance		Fumble Lobster Mumble
9	Lame Nap Sibling		Ambush Detonate Drone Infuse Hologram Telepathy Traitor Unicorn Weir	Nab Sheepish Snoop
10	Awesome Porn	Silo Vault	Cairn Cellular Download Flux Helm Interrogate Neural Odyssey Parchment Plasma Rendezvous Visor	Accomplice Clasp Impending Pigtail Tot
11	Closet	Freckle Fugitive	Genocide Hangar Lore Teal	Elf Scowl Slink

None of the learnable words in teacher talk are encountered in TV-talk as illustrated by the data in Tables 1 and 2 (see also Appendix A for an exhaustive list of the specific learnable word items for all 3 TV genres and for teacher talk). For instance, while the learnable word *fumble* is found in the teacher talk corpora, it is absent from the speech found in TV-talk. In addition, Table 5 shows that TV-talk offers more learnable words (n=122) than teacher talk (n=42). This indicates that there are 2.90 times as many words in TV-talk than in teacher talk. I hypothesize that this may be accounted for in part by the fact that the TV-corpus is roughly 3 times bigger than the teacher talk corpus, and the former is composed of 3 different corpora: sitcoms, dramas and sci-fi programs.

However, when the size of the TV-talk corpus is normalized (all the genres being combined to compile the TV-talk corpus and then adjusted to be comparable to the teacher talk corpus in terms of time of exposure; i.e., divided by 3), one can observe that the results are reversed: teacher talk now offers 1.33 more learnable words than TV-talk. This is the case despite the fact that the normalized TV-talk corpus is more than twice the size of the teacher talk corpus.

With regards to word repetition, contrary to initial expectations, TV-talk and teacher talk offer a similar pattern of learnable word repetitions. Assuming that learners will acquire a word faster if they encounter it more than 20 times, i.e. double the number of repetitions leading to potential acquisition (threshold established in Chapter 2), Table 7 lists all those learnable words by corpora (before the TV-talk corpus was normalized). When the size of the TV-talk corpus is normalized, 11.67 words are repeated 20 times or more. These results are not surprising given that teacher talk is catered to learners, unlike TV-talk, which is directed to an audience of fluent speakers/listeners and not produced

with vocabulary learning in mind. Based on these results, it appears that learners are likely to acquire the same amount of vocabulary whether they take 4 hours of ESL per week (the average time spend in the classroom in which the corpus was collected) or whether they watch 5.5 hours of television (which corresponds to the 50 hours of program collected for each genre of programs divided by 9 weeks, the length of the ESL course) for the same time period, i.e. a week.

Table 7 – Words Repeated 20 Times or More in Teacher Talk and TV-Talk

Teacher Talk		TV-Talk*	
Bout	Raft	Artefact	Bluffing
Dialogue	Sensor	Conduit	FBI
Lace	Telepathy	Directorate	Frakkin
Verb	Awesome	Insurgent	Krypter
Porridge	Download	Con	Malley
Pronunciation	Neural	Raptor	Oi
Braid	Odyssey	Psych	Psychiatrist
Newcomer	Plasma	Shuttle	Subspace
Trait	Porn	Yin	Tardis
Lobster	Vault	Roger	Tumor
Snoop	Lore	AI	ULD
Clasp	Teal	Baseship	

*List before normalization.

Figure 1 shows that most teacher talk learnable words are found in the BNC-5,000 to 7,000 word-levels shown in the blue, red and yellow parts of the T-talk bar on the right, that is 59.52% of the total learnable vocabulary for this corpus. Comparatively, there are no learnable words in these levels for TV-talk and the 3 first levels available in this corpus – BNC-8,000 to 10,000 – amount to 27.05% of the total number of TV-talk learnable words. There are in fact more learnable words – 30.33% – in the off-list section of the same corpus. In contrast, only 4.76% fall into this category for teacher talk.

Obviously, words in the lower, i.e. more frequent, BNC word-levels are more widely used in the English language than words in the higher, i.e. less frequent, BNC word-levels (including the last category: off-list words). Consequently, these results imply that the bulk of the words that can be acquired through teacher talk, which is found in the lower BNC word-levels, is useful in a wide variety of settings. On the other hand, the most important section of learnable words for TV-talk, observed in the off-list word-level, is likely to be of interest in much more specific contexts as they are less often used in the English language. Consequently, teacher talk seems to be more advantageous for intermediate learners while TV-talk proves more appropriate for the vocabulary acquisition of advanced learners.
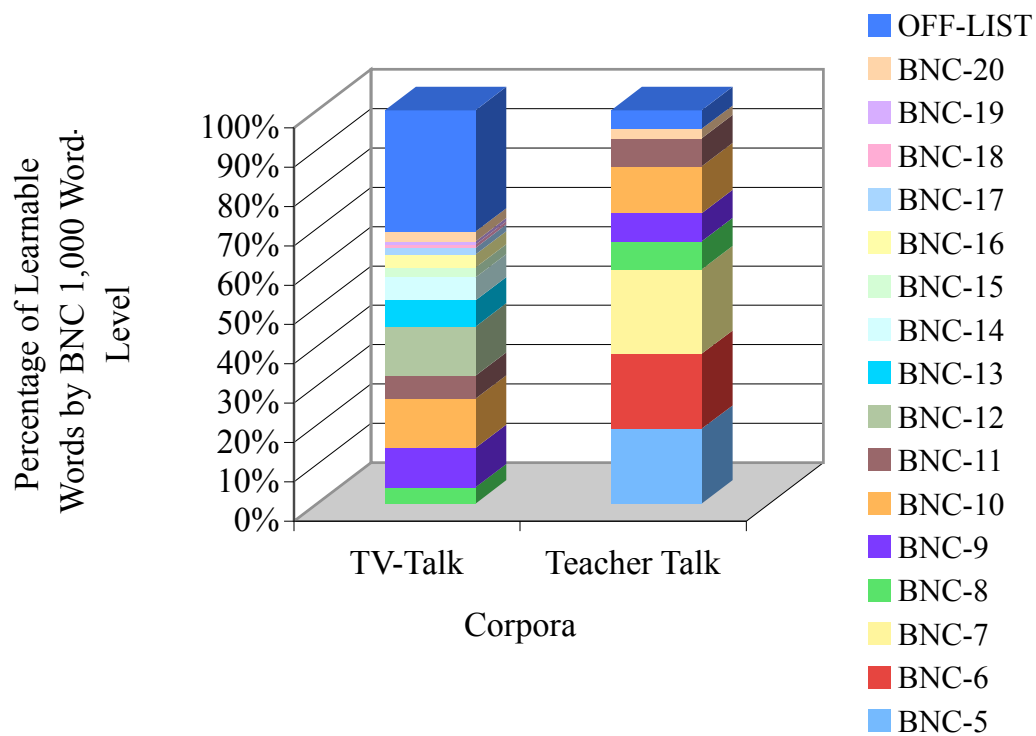
Figure 1 - Percentage of Learnable Words per BNC 1,000-
Word Level for TV-Talk and Teacher Talk

Via a Vocabprofile BNC-20 analysis of the drama corpus, one observes that learners need to know at least the 7,000 most frequent words before reaching the 98% known word coverage in this genre (see Table 3 for the complete BNC analysis), the threshold which allows them to acquire new words. These learnable words only appear at the BNC-8,000 word level. For sitcoms and sci-fi programs, the learnable words only appear at the BNC-9,000 word level (see Table 2 and Table 4 for the complete BNC analyses). Table 5 shows that teacher talk does not offer any learnable words in the 12,000 to 19,000 range of most-frequent words. Comparatively, TV-talk offers at least some words at all the levels where teacher talk does not.

Interestingly, as illustrated in Table 5 above, 57.37% of all learnable words are found within the BNC-8,000 to 14,000 word-levels for TV-talk. Adding the off-list learnable words to this vocabulary set brings this number up to 87.70%, that is, an additional 30.33%. As for teacher talk, the learners need to acquire knowledge of words up to and including the 7,000 word-level vocabulary to reach a comparable level of learnable words (59.52%). However, adding the off-list words to those words only brings up the potential for learning by an additional 4.76% for a total of 64.28% of the learnable vocabulary. These results therefore demonstrate that teacher talk offers more learnable words at lower, more frequent, BNC-word levels when compared to TV-talk, which in turn offers more learnable vocabulary at higher, less frequent, BNC-word levels. The bulk of the learnable words are concentrated in 3 word-levels for the teacher talk while it is more diverse in TV-talk. The latter also offers an important number of words from the off-list word-list, unlike teacher talk. Knowing where the learnable vocabulary is

concentrated can help the teachers select the activity that better suits the vocabulary learning goals of their students.

Semantically grouping learnable words together into various categories in each corpora (teacher talk and TV-talk) shows that not only do none of the words overlap in the two corpora, but that only one theme is common to both: dishonest activity/crime. While the most common categories in teacher talk are: school, body/physical appearance, body action, dishonest activity/crime and miscellaneous (see Table 8 for specific word classification), TV-talk is mostly characterized by the following categories: dishonest activity/crime, science, technology, sci-fi, sex, cursing and miscellaneous (see Table 9 for specific word classification). Some of the miscellaneous words of each corpora could belong to a category of the other corpus. However, in their respective corpora, they do not group with enough words to justify creating a category. For example, the word "freckle" in the TV-talk corpus could belong to the "body/physical appearance" teacher talk category and the word "mics" in the teacher talk corpus could belong to the "technology" TV-talk category.

Table 8 – Categories of Learnable Words in Teacher Talk

School	Body/Physical Appearance	Body Action	Dishonest Activity/Crime	Miscellaneous
Dialogue	Lace	Hiccup	Bout	Porridge
Verb	Pierce	Shiver	Standby	Newcomer
Noun	Hem	Blush	Accomplish	Lobster
Pronunciation	Trendy	Burp	Growl	Elf
Adjective	Braid	Mime	Teller	Mics
	Broach	Fumble	Perpetrate	
	Trait	Mumble	Nab	
	Pigtail		Sheepish	
	Tote		Snoop	
	Pantyhose		Accomplice	
			Clasp	
			Impending	
			Scowl	
			Slink	
			Swizzle	

Table 9 – Categories of Learnable Words in TV-Talk

Dishonest Activity/Crime	Science	Technology	Sci-Fi	Sex	Cursing	Miscellaneous	
Attorney	Dynamite	Cellular	Drone	Spank	Jackass	Raft	Sensei
Surveillance	Infuse	Download	Hologram	Porn	Frakkin'	Ranch	Slicer
Ambush	Neural	Malfunction	Telepathy	Syphilis	Frakking	Lame	Ump
Detonate	Plasma	Transceiver	Flux	Motel	Freaking	Nap	
Traitor	Lore	AI	Helm	Lingerie	Gosh	Sibling	
Interrogate	Artefact	Comm	Odyssey	Wingman	Jeez	Unicorn	
Vault	Balm	Deflector	Teal		Oi	Weir	
Visor	Directorate	Dradis	Airlock			Awesome	
Fugitive	DNA	Ipod	Conduit			Cairn	
Genocide	Pituitary	Nuke	Ensign			Parchment	
Hangar	Shale	ULD	Fissure			Rendezvous	
Autopsy	Magma		Wraith			Silo	
Gunshot	Aneurism		Android			Closet	
Insurgent	Chiropractic		Nebula			Freckle	
Disengage	Adipose		Raptor			Pumpkin	
Outpost	MRI		Shuttle			Nickel	
Con	Psychiatrist		Baseship(s)			VIP	
Roger	Tumor		Cloaking			Viper	
Cole			Hyperspace			Psych	
Bluffing			Interphasic			Rickshaw	
FBI			Krypter			Bagel	
Intel			Nova			Bongo	
Intercept			Phaser			Salsa	
			Resonator			Yin	
			Subspace			Busboy	
			Tardis			Fiancé	
			Teleport			Malley	

To sum up, teacher talk and TV-talk offer considerably different learnable words both in terms of specific items and vocabulary themes. In addition, they differ greatly in terms of word frequency, as measured by the BNC word-levels. Thus, the vocabulary learning goal determines which environment is most beneficial to the learners: teacher talk is more appropriate for intermediate learners, while TV-talk will benefit advanced learners. The learnable vocabulary in teacher talk is found mostly in the lower, more frequent, BNC word-levels while TV-talk provides most of its learnable words in low frequency BNC word-levels absent from teacher talk. For a similar time of exposure, contrary to what I expected, learnable vocabulary is not repeated more often in teacher than TV-talk. Lastly, teacher talk and TV-talk offer learnable vocabulary that significantly differs semantically, sharing only one common theme: "dishonest activity/crime".

In the case of TV-talk, how does the learnable vocabulary compare among the three genres considered in this study: sitcoms, dramas, and sci-fi programs?

The three genres of TV programs were compared to identify the vocabulary learning benefits of each. First, the cognitive load required for vocabulary acquisition was observed through the BNC word-level classification (see Tables 2 to 4 for the breakdown of the vocabulary per word level for each corpus) and it was shown that television watching for the purpose of learning vocabulary is only possible for advanced learners regardless of the genre of program – as per Nation's (2001) classification of learners defined in Chapter 2: all genres require the learners to know more than 5,000 words to reach the 98% known-word threshold.

The number of learnable words per BNC word-level was analyzed for each genre and the results compared to one another. The results show that most words are specific to one genre while very few are present in more than one – as discussed later, only four words overlap between genres. In fact, none of the learnable words are common to all three genres. Classifying the vocabulary by themes, I also classified their distribution across each genre, and then, the concentration of each genre per theme.

Analyzing the three corpora using the Vocab-profiler at Lextutor.ca allowed me to determine the 98% known-word coverage for each genre. Drama is the easiest genre to understand because it requires the learners to know only the 7,000 most-frequent words from the BNC list to reach the 98% threshold, while sitcoms and sci-fi programs tie at the BNC-8,000 word-level (see Tables 2 to 4 for the complete vocabulary profile analysis). However, according to Nation's (2001) vocabulary learning goals discussed in Chapter 2, even if dramas are slightly easier to understand than sitcoms and sci-fi programs, only advanced learners are likely to gain from watching any of these three genres of TV programs. This is because they all require the knowledge of at least 5,000 words, according to Nation, which corresponds to the vocabulary knowledge of advanced learners.

Sci-fi programs not only offer more learnable words than sitcoms and dramas – 68 words compared to 29 each for both sitcoms and dramas, as shown in Table 10, but they also offer the greatest concentration of learnable words in the lowest BNC word-levels. In fact, 34 of the 68 words (50%) available for uptake from this genre are from the BNC-9,000 to 12,000 word-levels. Thus, the contribution that sci-fi programs can make to learners' vocabulary acquisition is still substantial despite the fact that it also

introduces words that are only useful as part of the reality described in specific fiction realms. For instance, two of the learnable words from the sci-fi corpus – *phaser* and *resonator* – describe weapons that only exist in the reality of a single program: Star Trek – The Next Generation. These words do exist in the English language but not with the definition that is assigned to them in this sci-fi context. One swear word – *frak* – was also invented to replace *fuck* in the program Battlestar Galactica. Other words are used to describe sci-fi realities, e.g. *clone*, *drone*, *hologram*, but they are widely recognized and defined in dictionaries. By subtracting the three words that are program specific, the sci-fi corpus has in reality 65 learnable words to offer the learners. These results show that most of the learnable vocabulary available in sci-fi programs is in fact useful to the learners outside of the specific sci-fi program.

Table 10 – Number of Learnable Words per BNC Word-Level for Sitcoms, Dramas, Sci-

Fi Programs

BNC	Sitcoms	Dramas	Sci-Fi Programs
BNC-8	–	6	–
BNC-9	3	–	9
BNC-10	2	2	12
BNC-11	1	2	4
BNC-12	2	4	9
BNC-13	2	2	4
BNC-14	1	2	5
BNC-15	1	1	1
BNC-16	2	2	–
BNC-17	1	1	–
BNC-18	–	–	1
BNC-19	–	–	1
BNC-20	2	1	–
OFF-LIST	12	6	22
Total number of learnable words	29	29	68

Interestingly, most learnable words are specific to a particular genre (see Appendix B for an exhaustive list). More specifically, 26 learnable words are only available in sitcoms while the same is true for 25 words for dramas and 66 in sci-fi programs. Alternatively, very little learnable vocabulary – only 3.28% of the total number of TV learnable words – can be observed in more than one genre and none are present across all three genres. Only two words are observed in sitcoms and dramas (*con* and *AI*), 1 in sitcoms and sci-fi programs (*bluffing*) and 1 in dramas and sci-fi programs (*FBI*). These results imply that specific genres of TV programs will lead to specific word gains for the learners and that the genres are not interchangeable with regards to specific vocabulary learning goals.

Figure 2 illustrates the breakdown of learnable vocabulary when grouped by themes for each genre of programs. Interestingly, each genre appears to offer a unique breakdown of themes. In this analysis, all the learnable words were used, including the ones present in more than one program. Sitcoms offer a great amount of words in the "miscellaneous" category, but they provide a balance of learnable words across the five remaining themes. The only theme not present in this genre is "sci-fi" which, as expected, is also not present in dramas. Not surprisingly, this category of words is only learnable in the genre of TV program by the same name. Dramas, for their part, offer words in five different themes, one less than sitcom and sci-fi. Similar to sci-fi programs, dramas offer much more vocabulary in a number of categories – namely "crime", "science" and "miscellaneous" – while much less in others. The difference is that an important portion of the learnable words for sci-fi programs is associated with the technical vocabulary needed to describe the realities particular to this genre. The vast majority of the

remaining learnable vocabulary for this category relates to "crime", "science" and "technology". Note also that three of the four words that can be encountered in more than one genre are associated with "crime", while the last word refers to "technology". So, learners looking to improve their overall vocabulary knowledge would benefit more from sitcoms while learners wanting to acquire more vocabulary related to "crime" and "science" should opt for either dramas or sci-fi programs. Also, the latter genre offers the learners the added benefit of the "technology" vocabulary.

Sitcoms

Dramas
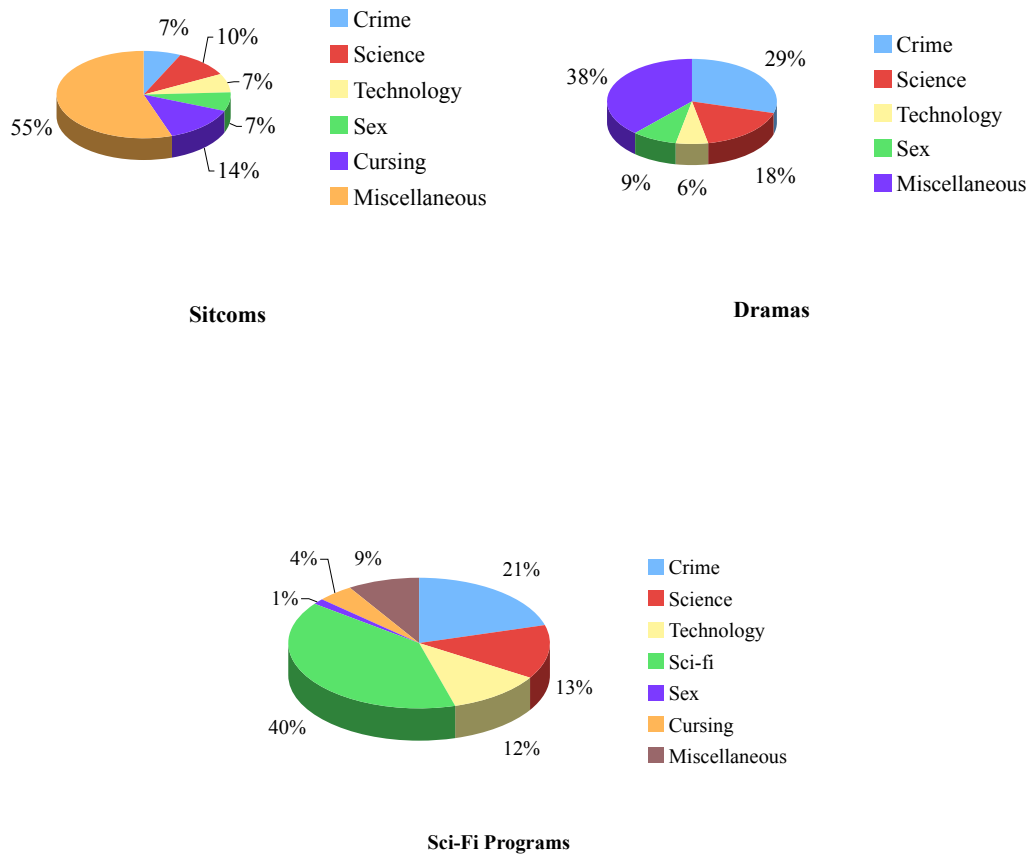
Sci-Fi Programs

Figure 2 – Distribution of Themes of Learnable Vocabulary by Genre

Table 11 shows that although dramas and sci-fi programs appear to offer a similar variety of word themes – with the exception of the sci-fi theme – sci-fi programs offer overall a higher number of learnable words for the same number of hours of exposure. This implies that learners should favor this genre of program if they want to maximize their vocabulary learning. On the other hand, if the learners want to acquire words from a wider variety of contexts, i.e. not specific to one theme, they should select sitcoms, since the most important number of learnable words comes from the "miscellaneous" section.

In brief, drama is slightly easier to understand than sitcoms and sci-fi programs. However, all three genres require the knowledge of a significant number of words and are only suitable for advanced learners. Sci-fi programs offer by far the greatest amount of learnable words while the other two genres tie at second place. Most learnable word items are specific to one genre and each genre of programs has a unique learnable vocabulary profile characterized by a number of dominant themes. Interestingly, sci-fi programs, aside from sci-fi related learnable words, offer possible vocabulary acquisition from a variety of themes that is similar to that of dramas. Sitcoms, on the other hand, are significantly different from the other two genres with the greatest concentration of its learnable words in the miscellaneous section.

Table 11 – Number of Learnable Words by Genre and Theme

| | | Themes | | | | | | | |
		Crime	Science	Technology	Sci-fi	Sex	Cursing	Miscellaneous	Total
Genres	**Sitcoms**	2	3	2	-	2	4	16	29
	Dramas	10	6	2	-	3	-	8	29
	Sci-fi Programs	14	9	8	27	1	3	6	68

What/how many words in teacher talk are more likely to become learnable when combined with TV-based homework?

One of the goals of this study was to determine how much more vocabulary and which words could be learnt by combining the teacher talk corpus with each TV-genre corpus (sitcom, drama and sci-fi). The assumption was that by combining the two activities, listening to teacher talk and listening to TV-talk, would increase sufficiently the number of repetition of the target learnable words (from less to more than 10 times), thus ensuring a context that would allow these new words to be learnt.

Based on the results obtained for the two previous questions, one may conclude that it is impossible to increase the learnable vocabulary by combining teacher talk with a television program. This is because the vocabulary demands of all TV genres are higher than that of teacher talk. In other words, teacher talk, which requires the learners to know 4,000 BNC word-levels, caters to the needs of intermediate learners who would be unable to understand the TV programs analyzed in the present research. As previously mentioned, learners require the knowledge of the 7,000 to 8,000 BNC word-levels to reach the 98% known-word coverage mark to be able to learn words from sitcoms, dramas and sci-fi programs. These findings make those programs only suitable for advanced learners, as per Nation's (2001) classification of English L2 learners. The cognitive loads involved would be too great for high-intermediate learners to understand the context, thus allowing them to decipher the meaning of unknown words. Inversely, if TV-talk is adequate for advanced learners, learnable words would likely not be found in a high-intermediate ESL class similar to the one in which the teacher talk was collected.

This question remains unanswered given the data at hand because the cognitive loads of

the TV-talk and teacher talk used in this research are different.

Chapter 5 – Discussion

This chapter discusses the results obtained and presented in the previous chapter in light of the literature on the subject. As was the case in Chapter 4, it is divided following the three research questions. First, teacher talk is compared to TV-talk in terms of word frequency (using the BNC classification) and actual word items (using a theme classification) are discussed. The three TV corpora are then compared amongst themselves using the same two types of classification used to address the first question. Finally, an explanation is given as to why the data at hand did not allow an answer to question 3, which aimed at determining the added potential benefits in terms of incidental vocabulary acquisition of combining exposure to teacher talk and TV-talk.

Teacher talk versus TV-talk

The analysis of the corpora suggests that teacher talk should be easier to understand than TV-talk. This holds true at least for the vocabulary environments under study in this research, where the learners need to know the most frequent 4,000 words (Nation, 2006) to reach 98% known word coverage of teacher talk, compared to the 7,000 to 8,000 most frequent words for TV-talk to attain the same coverage.

With respect to TV-talk, it was found that the 98% known-word coverage ranged from 7,000 to 8,000, depending on the genre of program. My results corroborate those of Webb and Rogers (2009) who found that the lexical demands of American programs requires the learners to know between 6,000 and 9,000 words to reach 98% known-word coverage. With regards to these vocabulary knowledge thresholds, the results obtained show a relatively more limited range of known-word coverage (7,000 to 8,000) than

those of Webb and Rogers. This could be explained by differing methods of data collection. As the two authors observed, results vary greatly from show to show and from episode to episode. Thus, having selected different programs and compiled larger corpora for the current research may explain the disparity in the results obtained.

As discussed in the previous chapter, teacher talk requires the knowledge of 4,000 words and TV-talk, 7,000 to 8,000 words to allow the learners to acquire new words from these sources of aural input. This difference in vocabulary demands between teacher talk and TV-talk can be explained by the fact that teacher talk caters to L2 learners and therefore is designed for learning purposes, while TV-talk is intended for L1 speakers and serves mostly as entertainment. Another explanation for the more manageable vocabulary load of teacher talk (meaning it is easier to understand than TV-talk) may be that the corpus was collected in a class of high-intermediate L2 learners of English.

As mentioned above, the cognitive load required to understand TV-talk is greater than that needed for teacher talk, yet TV-talk provides a similar amount of learnable words compared to the classroom environment. This means that for a similar number of hours of exposure, intermediate and advanced learners can acquire a similar amount of vocabulary, as long as the learners are exposed to the appropriate vocabulary learning environment (teacher talk or TV-talk). As demonstrated in Chapter 4 (Figure 1 and related discussion), after having been normalized to be comparable to teacher talk, TV-talk only offers 3.27% fewer learnable words than teacher talk. This percentage translates into a total of 1.33 additional learnable words in favor of teacher talk.

A difference of 1.33 learnable words in the current situation is negligible. Given that incidental vocabulary gains are very limited compared to focused vocabulary

instruction (Elley, 1989), previous research argued the opposite, claiming that a 1-word difference is significant. However, the corpora used in these studies were much more limited in size compared to the ones used in the present research. Previous research shows incidental vocabulary learning gains of 1 to 7 words when learners are exposed to 1,000 to 21,000 words (Day, Omura & Hiramatsu, 1991; Dupuy & Krashen, 1993; Horst, Cobb & Meara (1998). Comparatively, the current study suggests that the learners would be exposed to a much greater number of words – 111,812 words of teacher talk and more than 250,000 words of TV-talk. In such context, a difference of 1.33 learnable words in favor of teacher talk is unlikely to give a substantial advantage over TV-talk.

As reported in Chapter 4, learners encounter a similar number of learnable words when exposed to teacher talk as they would listening to a similar number of hours of TV-talk. However, the density of token words in TV-talk is more than double that of teacher talk in a comparable amount of time – 259,203 to 264,898 token words for TV-talk compared to 111,812 for teacher talk. In other words, students hear twice as much language in an hour of TV compared to an hour in the classroom exposed to teacher talk. These results are consistent with the fact that time spent in a classroom includes, but is not limited to, listening to teacher talk. The television programs selected are intended for the general English-speaking population and so the vocabulary is more varied and elaborate than vocabulary found in an L2 classroom, which tends to be simplified to accommodate the needs of the learners. In addition, the pace at which the speech is presented differs in the two environments. Teachers adapt their speech to their learners' need while TV-talk is presented at a faster pace than the one typically used in an L2 classroom.

These results are also consistent with Nation's (2001) findings, noting that the amount of token words to which learners need to be exposed in order to encounter new vocabulary increases with the size of the learners' vocabulary. In this particular case, learners exposed to teacher talk need to know 4,000 word families and be exposed to 111,812 running words, while they need to know at least 7,000 word families and be exposed to 250,000 running words of television programs to reach similar vocabulary gains. The cognitive loads of the two environments – teacher talk and TV-talk – are different, the latter being more demanding than the first. Thus, according to Nation's (2001) classification of learners, the teacher talk analyzed in this research would be better suited for intermediate learners and TV-talk for advanced learners.[8] Given that learners are exposed, for a similar amount of time, to the environment that suits their needs – teacher talk for intermediate learners and TV-talk for advanced learners – both groups of learners can potentially acquire a similar amount of vocabulary.

In other words, TV-talk can complement but not replace teacher talk when learners are involved in incidental vocabulary learning. As shown in Chapter 4, the two learnable-vocabulary environments do not overlap and, accordingly, they each offer different sets of learnable words. An important percentage of these words are found in the off-list section (30.33%) for TV-talk and the BNC-5,000 to 7,000 level (59.52%) for teacher talk. Comparatively, teacher talk only provides 4.76% learnable words from the off-list level while TV-talk offers no learnable words in the BNC-5,000 to 7,000 word-levels. This is not surprising since learners are expected to know these 5,000 to 7,000 words in order to reach the 98% known-word coverage that allows them to decipher the

---

[8] Intermediate learners need to know 3,000 to 4,000 words while advanced learners need to know at least 5,000 words.

meaning of unknown words. Furthermore, teacher talk offers no learnable words in the 12,000 to 19,000 range. On the other hand, TV-talk offers 57.37% of its learnable words between 8,000 and 14,000 (Table 5), which means it provides, unlike teacher talk, an important amount of words (66.39%) beyond the 11,000 mark. These results are consistent with the idea that classroom language appears to be simplified for ESL learners and that a more varied vocabulary is used for English L1 adult speakers (Horst, 2009). As the data analyzed here show (see Table 5), teacher talk offers more words in the lower, more frequent, BNC-word levels while TV-talk appears to spread over the entire spectrum. This entails that an important number of teacher talk words are likely to be encountered more often in everyday life than that of TV-talk. In brief, teacher talk learnable words do not overlap with those of TV-talk but are more likely to be used by the learners on a more regular basis outside of the classroom.

The usefulness of many of the learnable words available in TV-talk is not to be downplayed. As Nation (2001) points out, an educated L1 English speaker knows on average of 20,000 word families. One can then highlight the importance of acquiring the learnable vocabulary found beyond the BNC 11,000 word-level, where the majority of TV-talk learnable vocabulary is observed. In short, television programs are not only complementary to teacher talk for incidental vocabulary acquisition but rather provide an incidental learning environment for advanced learners who outgrew their L2 classroom.

Similar to the comparison in word levels, grouping the learnable vocabulary into categories shows that the themes of learnable words encountered in the two corpora are for the most part different, except for the "dishonest activities/crime" theme, which appears to be common to both types of speech, teacher talk and TV-talk. Otherwise,

teacher talk concentrates on "school", "body/physical appearance" and "body actions". Several words in teacher talk, more specifically the words belonging to the "school" theme such as *adjective*, *verb* and *pronunciation* are likely to be already known by the students. TV-talk would thus offer more learnable words than teacher-talk. Although TV-talk appears to address some of the shortcomings of the classroom environment in terms of learnable vocabulary, the gap that Horst (2010) found in teacher talk using the same corpus as in this study, i.e. words specific to "business topics", "government", and the "physical world" never appearing in the teacher talk, is unlikely to be filled by watching sitcoms dramas or sci-fi programs. As indicated earlier, the main topics of learnable words found in TV-talk focus on "science", "technology", "sci-fi", "sex" and "cursing". TV-talk, like teacher talk, does not offer a comprehensive list of useful learnable vocabulary.

Overall, TV-talk appears to provide more benefits to the learners. First, the school-related items (*dialogue*, *verb*, *noun*, *pronunciation* and *adjective*) are not only likely to already be known by high-intermediate learners of English, but are also likely to be of little use outside the classroom. TV-talk then provides six categories of learnable vocabulary compared to three for teacher talk (not counting the vast "miscellaneous" category). Furthermore, teacher talk learnable words also appear to be focused on people while TV-talk appears to offer a greater variety of subjects even if some, such as "sci-fi" vocabulary, may only be useful in very specific contexts, and others, such as "sex" and "cursing", may be inappropriate for formal encounters. To sum up, TV-talk addresses some of the shortcomings of teacher talk learnable vocabulary but the fact remains that it

offers some limitations of its own. In other words, neither of the speech environments provides a comprehensive list of themes for vocabulary acquisition.

In summary, teacher talk is easier to understand than TV-talk and each speech environment provides incidental vocabulary learning opportunities to a particular level of L2 English learners, respectively intermediate and advanced learners. Each type of talk offers their target learners a similar number of learnable words. However, these vocabulary items differ in the opportunities of use based on the BNC classification and on the theme classification. In the end, the vocabulary learning goals will determine which environment better suits the need of the learners. This will be further discussed in the next section as well as in Chapter 6.

Sitcoms, dramas & sci-fi programs

An examination of TV-talk and a comparison of the learnable vocabulary of the three genres revealed that dramas are slightly easier to understand than sitcoms and sci-fi programs. That is because learners need to know the first 7,000 BNC-word levels for dramas, while sitcoms and sci-fi programs demand the knowledge of the first 8,000 BNC-word levels (Table 5) to reach the 98% known-word coverage necessary for comprehension and for interpreting the meaning of unknown words. According to Nation's (2001) vocabulary learning goals, the results obtained suggest that the three TV genres investigated in this research are only beneficial for advanced learners because they require the knowledge of more than 5,000 words. This is not surprising given that these shows are intended for adult native English-speaking viewers.

The findings obtained by Webb and Rodgers (2009) are slightly different than the ones reported in Chapter 4 in that the two authors found that the knowledge of 6,000

words was necessary to reach 98% known-word coverage for dramas and sitcoms, and as much as 9,000 words for sci-fi programs. However, as the authors point out, there are great variations from episode to episode. Since the size of Webb and Rodgers' corpora was very limited and unevenly compiled (see Chapter 2 for the data they reported), it is possible that those limitations affected the results. Despite these shortcomings, it is interesting to note that Webb and Rogers also found that all three genres are only suitable for advanced learners.

Different types of programs, such as programs catering to a younger audience, may be more appropriate to learners of lesser proficiency levels. Intuitively, I would have assumed that programs aimed at children would have offered a learning environment adequate for less advanced learners. However, research point to the contrary. Nation (2006) and Webb and Rogers (2009) have shown that television programs for younger audiences such as the movie *Shrek* and the TV shows *Fraggle Rock*, *Mr. Rogers* and *Sesame Street* also provide a listening environment appropriate for advanced learners' incidental vocabulary acquisition. However, further research is necessary to determine if these results can be generalized to youth television as a whole or if some youth programs could cater to the needs of less proficient learners. As well, more research is needed to determine if other genre of shows, such as documentary and reality TV, would result in different findings.

Vanderplank (1993) observed that based on students' interests, the learners favor sitcoms because they enjoy laughing while learning. Laughing in the right places for the right reasons may be encouraging for the learners and consequently draw them to favor this genre more than others. However, the current research shows that our intuitions do

not always lead us towards the most favorable environments for vocabulary learning. As Chapter 4 indicates, of all three genres studied, sitcoms are the least desirable choice in providing the environment for incidental vocabulary acquisition. As mentioned above, sitcoms require a more elaborate vocabulary knowledge than dramas do. Second, sci-fi programs, which have the same vocabulary demands as sitcoms, offer more potentially learnable words than sitcoms do. Thus, teachers basing their selection of TV programs solely on students' interests may not provide their learners with the optimal environment for incidental vocabulary acquisition.

In terms of the amount of learnable vocabulary available, Chapter 4 reports that sci-fi programs offer by far the most number of words – 68, compared to 29 for both dramas and sitcoms. A possible explanation for the disparity in the number of learnable words between sci-fi programs and the other genres is the fantasy nature of the shows. In sitcoms and dramas, the terms describing the settings and many aspects of the stories likely refer to concepts that are common and that can be found in the lower BNC-levels, thus are already known. Comparatively, those terms are likely to be repeated as often in sci-fi programs but, because they describe realities that are imaginary, e.g. *cloaking*, *hyperspace* and *teleport*, many are found in the off-list BNC word-level and are thus, learnable rather than learnt. Interestingly, 27 of the 68 sci-fi learnable words are directly related to science fiction, as discussed earlier. The fact remains that not counting sci-fi specific learnable words, sci-fi programs still offer 41 learnable words useful in and outside the fiction realm, that is 12 more than what dramas and sitcoms each provide. Therefore, learners who watch television to improve their vocabulary should opt for sci-fi programs in order to maximize their vocabulary uptake.

Sci-fi programs also offer the greatest concentration of learnable words in the lower BNC word-levels. 34 of the 68 words (50%) are found between the 9,000 and 12,000 BNC word-levels, as illustrated in Chapter 4. A possible explanation for these results may be that the sci-fi program writers have tried to target a particular audience. In the end, regardless of the motivation, the stylistic decision made to produce these shows did result in the use of a less elaborate vocabulary. Given Nation's (2001) classification of word-levels, it appears that the most useful learnable words available in TV-talk are found in greatest concentration in the sci-fi programs. Once again, this genre of TV programs appears to be the most beneficial to learners in terms of learnable vocabulary.

Contrary to my initial prediction, very few of the learnable words available in sci-fi programs are specific to the reality created in individual programs. Aside from *phaser* and *resonator* from *Star Trek* and *frak* from *Battlestar Galactica*, the results show that most learnable words present in the sci-fi programs are useful in a variety of contexts, including, but not limited to, sci-fi programs. As discussed in Chapter 4, 27 of these words do belong to the sci-fi category and refer to concepts that are widely recognized, e.g. *clone*, *drone* and *hologram* (see Table 9 for the complete list). Accordingly, these words are known in the sci-fi universe in general not just as part of the setting of a particular story. Of the 41 remaining sci-fi learnable words, 14 relate to crime, 8 are about technology and 9, science. In absolute terms, sci-fi programs offer more learnable words in these 3 categories than the other two genres each offer in total. Although the sci-fi specific vocabulary is not found in the other genres, 5 of the 6 remaining categories of words are present in all three genres. This suggests that the words related to these 5 themes are worth learning. On the other hand, the use of a few learnable words such as

those found in the "sex" category (present in all three genres) or in the "cursing" category (present in sitcoms and sci-fi programs), is likely to be somewhat less useful than other learnable words as their use is appropriate in much fewer contexts. Overall, in trying to maximize incidental vocabulary acquisition by watching television, sci-fi programs consistently appear to be the better choice.

Most learnable words offered by the TV corpora appear to be specific to one genre (see Appendix B for an exhaustive list). 26 of these words are only found in sitcoms, 25 solely in dramas and 66 in sci-fi programs only. As reported in Chapter 4, only 4 TV learnable words (2 in sitcoms and dramas, 1 in sitcoms and sci-fi programs and 1 in dramas and sci-fi programs) can be observed in more than one genre. None were found to carry across all three genres. These results are consistent with those of MacFadden, Barren and Horst (2009) who found that learners who acquire the knowledge of a 689-word Television Word List (TWL) only increase their comprehension by 1-2% depending on the particular show. The authors created this list using range, i.e., words that appeared 7 times or more in at least 5 of the 10 shows studied, and looked at the vocabulary beyond the 2,000 level mark. When running their TWL through the Lextutor Vocabprofiler, I observed that the first 7,000 words cover 97.03% of the list and that an extra 1,000 words bring this total to 97.69%, which leaves less than 3% over the 7,000 or 8,000 BNC word-levels. In absolute terms this means that only 15.92 to 20.46 TWL words are present across different shows. In other words, very few TWL words beyond the 7,000 to 8,000 mark found across different genres. The fact that the current study used 10 repetitions (and not 7) renders the possibility of observing learnable words across a variety of genres even less likely. My results suggest that

specific genres of TV programs will lead to specific word gains. In turn, teachers must carefully identify the vocabulary learning goals to appropriately select the television programs that suit their learners' needs.

When the learnable words are broken down into categories by themes, it was also observed, in Chapter 4, that some genres offer a greater concentration of learnable words in a number of these themes. Sitcoms offer a large portion (55%) of its learnable words in the "miscellaneous" category. A possible explanation for these results is that sitcoms' stories are driven by humor and so do not need to be developed to the same depth that dramas and sci-fi programs do to reach their goal. This finding is consistent with the fact that the authors often select words for comedic effect that are not commonly used otherwise. The remainder of the learnable vocabulary is evenly distributed across the themes that it offers.

In contrast, the previous chapter illustrates that dramas, while also offering an important amount of learnable vocabulary in the "miscellaneous" category, provide much more vocabulary in the "crime" and "science" categories. Comparatively, sci-fi offers the greatest concentration of vocabulary in the "crime", "science" and "technology" categories (interestingly, the few words encountered in more than one genre are associated with "crime" and "technology"). A possible explanation for these results is that dramas and sci-fi programs' stories are developed with more depth than sitcoms' stories, which are driven by humor and so do not need to be as developed to reach their goal. Thus, the learnable vocabulary is less diverse in dramas and sci-fi programs compared to sitcoms as the words used are directly related to the subject matters developed in the television programs.

In short, learners looking to improve their overall vocabulary knowledge should opt for sci-fi programs. Even though the learnable vocabulary this genre offers is more focused on certain themes, namely "sci-fi", "crime", "science" and "technology", it does offer the most important number of words, by far. However, learners looking to improve their vocabulary in a variety of themes might be more interested in sitcoms, which offer vocabulary that is more evenly distributed across the different themes. Learners wanting to learn vocabulary associated with "crime" and "science" but who do not particularly enjoy science fiction may find dramas to be a suitable alternative.

Combining teacher talk and TV-talk

One of this study's goals was to find out which and how many words in teacher talk can become learnable when combined with TV-based homework. The results reported in Chapter 4 indicate that the three TV genres considered would only provide learnable vocabulary for advanced learners while teacher talk would do so more efficiently for intermediate learners. Based on these results, it was concluded that combining the two corpora would not be appropriate, as it the cognitive loads of the corpora are not compatible. The TV programs used in this research requires learners to know 7,000 to 8,000 words to reach the 98% known-word coverage necessary to acquire new vocabulary. As previously mentioned, it is assumed that intermediate learners know between 3,000 and 4,000 words, a number that is not sufficient for incidental vocabulary acquisition using the TV shows selected in this study. Watching these shows would thus be of little benefit to these learners who would be unable to interpret the meaning of unknown words, which could then potentially lead to vocabulary acquisition.

Future research may find television programs that could be combined with teacher talk to determine the answer to this question. However, the results of existing research have not been optimistic: even television aimed at a younger audience still appears to require the knowledge of a substantial number of words so that it reach the 98% known-word coverage. As Nation (2006) found, the children's movie *Shrek* requires the learners to know 7,000 words to reach this 98% threshold. Similarly, Webb and Rodgers (2009) found that learners needed to know 5,000 words to reach the same coverage with children's programs. Their findings are consistent with the knowledge of the target audience; as Nation (2001) points out, 5-year olds know on average about 5,000 words. Even though it is possible that future research identifies television programs with a cognitive load that is appropriate for beginners or intermediate learners, it is rather improbable.

Similarly, future research could collect and analyze teacher talk in an advanced ESL course to determine whether the teacher talk to which these learners are exposed could be compatible with the television programs under study. If the teacher talk of advanced ESL courses proves to require knowledge of at least 5,000 BNC word-levels, it would then be possible to find out which and how many learnable words are available to advanced ESL learners.

Nonetheless, given that teacher talk provides a limited number of learnable words at the higher, low frequency, BNC word-levels, it seems important to complement classroom teaching with television viewing, as this activity appears to fill some of the vocabulary gaps found in teacher talk for the advanced student. Thus, ESL teachers should keep in mind that contrary to beginners and intermediate learners, advanced

learners appear to have a lot to gain in terms of incidental vocabulary acquisition by incorporating television watching of sitcoms, dramas and especially sci-fi programs.

Chapter 6 – Implications and Conclusions

The general goal of this research was to identify the benefits of adding 50 hours of television viewing to an ESL language course. It ascertained the incidental vocabulary learning opportunities provided by the ESL learners' exposure to sitcoms, dramas and sci-fi programs, and compared these opportunities to those provided by teacher talk. In addition, I hoped to establish the added benefits of combining exposure to teacher talk and TV-talk as a tool to enhance the teaching of vocabulary. As Horst (2010) points out, the classroom suffers some shortcomings with regards to incidental vocabulary acquisition, seeing that certain topics and related vocabulary such as government or the physical world, are absent from teacher talk. Television programs were thought to possibly offer an alternative to alleviate these shortcomings. As such, this research has implications both at the pedagogical level to support teachers in their selection of additional materials for their students, but also for autodidact learners who wish to maximize their vocabulary learning opportunities by watching television programs.

The current study compared TV-talk to teacher talk, contrasted specific genres of television programs amongst one another and addressed some of the comparability issues of other studies by using television corpora of similar sizes. The teacher talk corpus was however smaller than the corpora of television programs but mathematical measures were taken in the analysis to ensure comparability with the television genre corpora. The language course used for this study spanned over 32 hours of teacher talk and amounted to 111,812 token words. The 3 television genres under study each covered 50 hours of episodes and amounted to a total of 259,203 words for sitcoms, 264,898 for dramas and 250,067 for sci-fi programs. Even though the size of each of the TV-talk corpora in terms

of token words was more than double that of teacher talk, the resulting potential vocabulary gains of teacher talk were similar to those of the normalized TV-talk.

I acknowledge the limitations of comparing corpora of TV-talk and teacher talk compiled using duration rather than number of token words as a measure of length. The current research compared the availability of learnable words in teacher talk and TV-talk given a similar time of exposure to speech. However, as Webb and Rodgers (2009) point out, the size of corpora, measured using the number words, can significantly affect the results. Thus, comparing corpora containing similar amounts of token words could yield different results.

Various factors should be taken into account when choosing to expose learners to teacher talk or TV-talk. In terms of sheer amount of word exposure, the analysis indicates that teacher talk is more efficient at providing learnable vocabulary. In other words, learners in a classroom environment are exposed to about half the number of words they would encounter were they to watch TV programs but the amount of learnable vocabulary the learners are expected to acquire is the same for both environments. On the other hand, the learners spend approximately the same amount of time listening to TV-talk as they would for teacher talk in order to be exposed to twice as many running words as they would be in the classroom. What this means for teachers is that, the amount of running words may not be the primary factor when making the pedagogical decision of exposing learners to teacher talk or TV-talk. In addition, it appears that teacher talk and TV-talk caters to the needs of students with different levels of proficiency. The findings suggest that teacher talk is more suitable for intermediate learners while TV-talk caters to

the needs of advanced students, as per Nation's (2001) classification of proficiency levels.

The learnable vocabulary offered by teacher talk is different from that of TV-talk. Some overlap was observed in the BNC word-levels 9,000 through 11,000; however, none of the learnable items identified are present in both corpora. In addition, teacher talk provides more learnable words in the lower BNC word-levels. When grouping the learnable words into themes, important differences in TV-talk and teacher talk were noticed. For teacher-talk, a few of the learnable words pertain to "school", while the rest refer to "body/physical appearance", "body actions" and "dishonest activities". The majority of TV-talk learnable words refer to "crime", "science", and "technology", while the rest pertain to "sci-fi", "sex" and "cursing". Aside from the "dishonest activities" and "crime" categories, the breakdown of themes for the learnable vocabulary of teacher talk and TV-talk is significantly different. The finding that the learnable vocabulary profiles of teacher talk and TV-talk are inherently different will aid teachers to better plan their lessons. Specifically, gaining awareness of what kind of vocabulary (classified by BNC word-levels or by themes) is present in the environments studied herein will allow teachers to make more informed choices on classroom activities suited to the needs of their students.

Taking a closer look at the different components of the TV-talk corpus, I found that although all genres would only benefit advanced learners, dramas are slightly easier to understand. Interestingly, the great majority of the learnable vocabulary is specific to one genre and only a handful of words can be observed across genres. These data suggest that each genre provides learners with specific benefits. Overall, sci-fi programs appear to

be the more advantageous choice when it comes to incidental vocabulary learning, both because of the sheer number of learnable words available, and also for the variety of vocabulary themes it offers. Dramas' breakdown of vocabulary themes is the one that resembles the most that of sci-fi programs, but because it offers fewer learnable words, only the learners' personal tastes can justify choosing this genre over sci-fi programs. Sitcoms may be more attractive to learners looking to challenge themselves rather than learners who want to improve their vocabulary. This is because this genre appears to provide very specific vocabulary (more than half of the learnable vocabulary cannot be grouped into a particular theme). The fact that over 40% of the learnable vocabulary available in this genre is from the off-list section means that learners' future use of these words is likely to be limited.

I do not suggest that when it comes to acquiring vocabulary, exposure to meaning-focused speech (be it teacher talk or TV-talk) should replace more direct approaches in the classroom. Horst (2010) points out that teacher talk offers little support for incidental vocabulary gains, similar to what was found in my TV-talk corpus. As Elley's (1989) study shows and Nation (2001) advocates, focused attention to vocabulary is much more desirable and much more efficient than incidental vocabulary acquisition. However, given that learners acquire vocabulary in a variety of ways, teachers should make informed pedagogical decisions that include the knowledge of the learnable vocabulary available for uptake in the materials they use.

Hoping to maximize the vocabulary learning available, I initially intended to combine the teacher talk corpus and the various TV-talk corpora to determine the resulting increase in learnable vocabulary. However, after finding the vocabulary

knowledge required of each corpus to reach the 98% known-word coverage, I observed that the teacher talk and the TV-talk were not compatible. It was found that the cognitive loads of each type of speech were different and that, in fact, teacher talk caters to intermediate learners, while TV-talk caters to advanced learners, thus making it impossible to combine the corpora. In other words, while teacher talk is appropriate for intermediate learners, TV-talk is too difficult for the same learners.

ESL teaching should consist of a variety of learning activities and, based on the results of this study, a case can be made for learners' exposure to both teacher talk and TV-talk. Even if other methods, such as focused attention on vocabulary, yield better results (Elley, 1989), incidental vocabulary learning remains part of the process. Teachers should be aware of the benefits and limitations of exposing learners to teacher speech. As the results obtained show, this type of input is better suited for incidental vocabulary acquisition by intermediate learners. In addition, advanced learners should be encouraged to complement their learning by watching television since, at that point in their learning, teacher talk has little left to offer. At first, dramas should be favored because this genre is slightly less demanding than sitcoms and sci-fi programs. Then, learners should be encouraged to watch sci-fi programs not only for the sheer number of learnable words available in this genre, but because, compared to sitcoms, they offer vocabulary pertaining to a greater variety of themes. Learners' interests in a particular genre of programs remain an important consideration for teachers when selecting television viewing content for their students. However, as a consequence of my findings, teachers' pedagogical decisions may also be guided by the potential incidental vocabulary benefits of the various television genres.

Aural incidental vocabulary acquisition is a fairly new area of study in applied linguistics and little has been done with regards to TV-based vocabulary learning environments. Much more research is needed on the subject to determine the true value of television watching for ESL teaching and learning purposes. Other types of programs such as documentaries and reality shows need to be explored. Just as Webb and Rodgers (2009) did with British and American English, television entertainment of different English speaking regions needs to be compared. Webb and Rogers, and MacFadden, Barrett and Horst (2009) also noticed that older shows appear to offer a somewhat different vocabulary profile. This is consistent with the fact that language is constantly evolving. Thus, ongoing research may be warranted to take note of the changes brought by time, so that teachers have up-to-date data to make pedagogical decisions with regards to television watching.

However, before research can become truly meaningful in this area, the threshold for incidental vocabulary learning from aural input should be established. To date, I have only been able to speculate on what the potential for such input is and the implications that were deduced from the results remain hypothetical. Nonetheless, my results show that television programs fill part of the gap in incidental vocabulary acquisition left by teacher talk.

References

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics, 24*, 425-438.

Bonk, W. J. (2000). Second Language Lexical Knowledge and Listening Comprehension. *International Journal of Listening, 14*, 14-31.

Brown, R., Waring, R., & Donkaewbua. S. (2008). Incidental vocabulary acquisition through reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*(2), 136-163.

Carver, R. P. (1994). Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction. *Journal of reading Behavior, 26*, 413-437.

Cobb, T. (2010). The compleat lexical tutor (version 6.1) [Computer software]. Retrieved from http://www.lextutor.ca

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language, 25*, 1-18.

Day, R. R., Omura, C., & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language, 7*(2), 541-551.

Deighton, L. (1959). *Vocabulary development in the classroom*. New York: Colombia University Press.

Dupuy, B., & Krashen, S. (1993). Incidental vocabulary acquisition in French as a second language. *Applied Language Learning, 4*(1-2), 55-63.

Duquette, L., & Painchaud, G. (1996). A comparison of vocabulary acquisition in audio and video contexts. *Canadian Modern Language Review, 53*, 143-172.

D'Ydewalle, G., & Van de Poel, M. (1999). Incidental foreign-language acquisition by children watching subtitled television programs. *Journal of Psycholinguistic Research, 28*(3), 227-244.

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 24*, 174-187.

Ellis, R. (1995). Modified oral input and the acquisition of word meanings. *Applied Linguistics, 16*(4), 409-441.

Ewers, C. A., & Brownson, S. M. (1999). Kindergarteners vocabulary acquisition as a function of active vs. passive storybook reading, prior vocabulary, and working memory. *Journal of Reading Psychology, 20*(1), 11-20.

Fraser, C. A. (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition, 21*(2), 225-241.

Gitlin, T. (2000). *Inside Prime Time*. California: University of California Press.

Goodman, J. C., Dale, P. S., & Li, P. (2008) Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language, 35*(3), 515-531.

Haastrup, K., Albrechtsen, D., & Henriksen. (2004). Lexical inferencing processes in L1 and L2: same or different? *Angles on the English-Speaking World, 4*(1-2), 111-128.

Henriksen, B., Albrechtsen, D., & Haastrup, K. (2004). The relationship between vocabulary size and reading comprehension in the L2. *Angles on The English-Speaking World, 4*(1-2), 129-140.

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*(2), 689-696.

Horst, M. (2010). How well does teacher talk support incidental vocabulary acquisition? *Reading in Foreign Language, 22*(1), 161-180.

Horst, M. (2009). Revisiting classrooms as lexical environments. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical processing in second language learners: Papers and perspectives in honour of Paul Meara* (pp. 53-66). Clevedon: Multilingual Matters.

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

Horst, M., & Meara, P. (1999.) Test of a model for predicting second language lexical growth through reading. *The Canadian Modern Language Review, 56*(2), 308-328.

Hu, M., & Nation, I.S.P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403-430.

Huckin, T., & Coady, J. (1999). Incidental Vocabulary Acquisition in a Second Language. *Studies in Second Language, 21*(2), 181-193.

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

Jones, L. C., & Plass, J. L. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal, 86*(4), 564-561.

Jopp, D. S., & Hertzog, C. (2010). Assessing adult leisure activities: An extension of self-reporting activity questionnaire. *Psychological Assessment, 22*(1), 108-120.

Koolstra, C. M., & Beentjes, J. W. J. (1999). Children's vocabulary acquisition in a foreign language through watching subtitled television programs at home. *Educational Technology Research and Development, 47*(1), 51-60.

Koskinen, P. S., Knable, J. E., Markham, P. L., Jensema, C. J., & Kane, K. W. (1996). Captioned television and the vocabulary acquisition of adult second language correctional facility residents. *Journal of Educational Technology Systems, 24*(4), 359-373.

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. *Modern Language Journal, 73*, 440-464.

Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316-323). Clevedon, UK: Multilingual Matters.

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

Lévesque, M. (in press). How large a vocabulary is needed to understand sitcoms and dramas: A TV corpus analysis. *Concordia Working Papers in Applied Linguistics*, 4.

MacFadden, K., Barrett, K., & Horst, M. (2009). What's in a television word list? A corpus-informed investigation. *Concordia Working Papers in Applied Linguistics 2*(1), 78-98.

Meara, P. (1993). Tintin and the World Service: a look at lexical environments. *IATEFL Annual Conference Report*, 32-37.

Meara, P. Lightbown, P. M., & Halter, R. H. (1997). Classrooms as lexical environments. *Language Teaching Research, 1*(1), 28-47.

Mueller, G. (1980). Visual contextual cues and listening comprehension: An experiment. *Modern Language Journal, 64,* 335-340.

Na, L., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context. *RELC Journal, 16*(1), 33-42.

Nagy, W., Anderson, R. C., & Herman, P. (1987). Learning word meanings from context during normal reading, *American Educational Research Journal, 24*, 237-270.

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

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

Nippold, M. A., Duthie, J. K., & Larsen, J. (2005). Literacy as a leisure activity: Free-time preferences of older children and young adolescents. *Language, Speech, and Hearing Services in Schools (LSHSS), 36*(2), 93-102.

Oetting, J. B., Rice, M. L., & Swank, L. K. (1995). Quick incidental learning (QUIL) of words by school-age children with and without SLI. *Journal of Speech and Hearing Research, 38*, 434-445.

Paribakht, T. S., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition: An introspective study of lexical inferencing. *Studies in Second Language Acquisition, 21*(2), 194-224.

Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics, 18*(1), 41-58.

Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew Effect? *Journal of Education Psychology, 94*(1), 23-33.

Rice, M. L., & Woodsmall, L. (1988). Lessons from television: Children's word learning when viewing. *Child Development, 59*, 420-429.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition, 21*, 589-619.

Senechal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *Journal of Child Language, 24*(1), 123-138.

Schonell, F. J., Meddleton, I. G., & Shaw, B. A. (1956). *A study of the oral vocabulary of adults*. Brisbane: University of Queensland Press.

Sykes, P. (2010). *Top 100 Sci-Fi TV Shows List*. Retrieved July 29, 2010 from http://home.austarnet.com.au/petersykes/topscifi/lists_tv.html.

Tyson, M. (1990). The effect of media on listening comprehension. Research notes. *The Oregon Teachers of English to Speakers of Other Languages Journal, 11*, 85-93.

Uchikoshi, Y. (2006). English vocabulary development in bilingual kindergarteners: What are the best predictors? *Bilingualism: Language and Cognition, 9*(1), 33-49.

Van Zeeland, H., & Schmitt, N. (Advanced Access published December 18, 2012). *Lexical coverage and L1 and L2 listening comprehension: The same or different from reading comprehension? Applied Linguistics,* doi:10.1093/applin/ams074.

Vanderplank, R. (1993). A very verbal medium: Language learning through closed captions. *TESOL Journal, 3*(1), 10-14.

Vidal, K. (2002). Academic listening: A source of vocabulary acquisition? *Applied Linguistics, 24*(1), 56-89.

Webb, S., & Rodgers, M. P. H. (2009). Vocabulary demands of television programs. *Language Learning, 59*(2), 335-366.

Wesche, B., & Paribakht. (2010). *Lexical inferencing in a first and second language: Cross-linguistic dimensions,* Buffalo, NY: Multilingual Matters.

West. M. (1953). The technique of reading aloud to a class. *English Language Teaching Journal, 8*(1), 21-24.

Wode, H. (1999). Incidental vocabulary acquisition in the foreign language classroom. *Studies in Second Language Acquisition, 21*(2), 243-258.

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

Number of Appearances of Learnable Words by BNC 1000-Word Frequency Increments
for Sitcoms, Dramas, Sci-fi Programs and Teacher Talk

Words	BNC	Sitcoms	Dramas	Sci-Fi Programs	Teacher Talk
bout	BNC-5				28
dialogue	BNC-5				47
hiccup	BNC-5				12
Korea*	BNC-5				16
lace	BNC-5				28
pierce	BNC-5				16
shiver	BNC-5				13
standby	BNC-5				12
verb	BNC-5				35
accomplish	BNC-6				17
growl	BNC-6				10
hem	BNC-6				10
noun	BNC-6				13
porridge	BNC-6				23
pronunciation	BNC-6				24
teller	BNC-6				10
trendy	BNC-6				12
adjective	BNC-7				14
blush	BNC-7				10
braid	BNC-7				27
broach	BNC-7				11
burp	BNC-7				11
mime	BNC-7				10
newcomer	BNC-7				22
perpetrate	BNC-7				11
trait	BNC-7				20
attorney	BNC-8			14	
Cuba*	BNC-8			12	
dynamite	BNC-8			10	
fumble	BNC-8				10
lobster	BNC-8				26
mumble	BNC-8				14
raft	BNC-8			20	
ranch	BNC-8			17	
spank	BNC-8			10	
surveillance	BNC-8			11	

Word	BNC				
ambush	BNC-9			10	
detonate	BNC-9			12	
drone	BNC-9			10	
infuse	BNC-9			10	
hologram	BNC-9			12	
lame	BNC-9	16			
nab	BNC-9				13
nap	BNC-9	15			
sibling	BNC-9	10			
sheepish	BNC-9				12
snoop	BNC-9				22
telepathy	BNC-9			30	
traitor	BNC-9			19	
unicorn	BNC-9			10	
weir	BNC-9			11	
accomplice	BNC-10				17
awesome	BNC-10	62			
cairn	BNC-10			14	
clasp	BNC-10				23
cellular	BNC-10			16	
download	BNC-10			24	
flux	BNC-10			10	
helm	BNC-10			11	
impending	BNC-10				13
interrogate	BNC-10			19	
neural	BNC-10			25	
odyssey	BNC-10			53	
parchment	BNC-10			12	
pigtail	BNC-10				11
plasma	BNC-10			33	
porn	BNC-10	24			
rendezvous	BNC-10			19	
silo	BNC-10		18		
Thanksgiving*	BNC-10	36			
tote	BNC-10				14
vault	BNC-10		36		
visor	BNC-10			13	
closet	BNC-11	11			
elf	BNC-11				11
freckle	BNC-11		13		
fugitive	BNC-11		10		
genocide	BNC-11			10	

hangar	BNC-11			11	
lore	BNC-11			30	
scowl	BNC-11				10
slink	BNC-11				10
teal	BNC-11			60	
airlock	BNC-12			14	
artefact	BNC-12			40	
Atlantis*	BNC-12			20	
autopsy	BNC-12		17		
balm	BNC-12	15			
conduit	BNC-12			21	
directorate	BNC-12		39		
DNA	BNC-12			13	
ensign	BNC-12			17	
fissure	BNC-12			16	
gunshot	BNC-12		12		
insurgent	BNC-12			21	
pituitary	BNC-12			11	
pumpkin	BNC-12	19			
syphilis	BNC-12		17		
wraith	BNC-12			10	
disengage	BNC-13			11	
nickel	BNC-13	14			
malfunction	BNC-13			11	
motel	BNC-13		18		
outpost	BNC-13			15	
Panama*	BNC-13		21		
shale	BNC-13		14		
VIP	BNC-13	10			
viper	BNC-13			13	
android	BNC-14			15	
ba*	BNC-14			39	
con	BNC-14	16	21		
lingerie	BNC-14			15	
magma	BNC-14			11	
nebula	BNC-14			10	
raptor	BNC-14			31	
transceiver	BNC-14		10		
psych	BNC-15		23		
rickshaw	BNC-15	10			
shuttle	BNC-15			25	
aneurism	BNC-16		10		

bagel	BNC-16	14			
Bonnie*	BNC-16	12			
bongo	BNC-16		15		
chiropractic	BNC-16	12			
Gael*	BNC-17	11			
salsa	BNC-17	15			
yin	BNC-17		25		
adipose	BNC-18			19	
Jerry*	BNC-19	196			
roger	BNC-19			20	
busboy	BNC-20	12			
cole	BNC-20		14		
jackass	BNC-20	10			
swizzle	BNC-20				12
AI	OFF-LIST	13	91		
Anna*	OFF-LIST			10	
baseship(s)	OFF-LIST			43	
bluffing	OFF-LIST	10		12	
cloaking	OFF-LIST			14	
comm	OFF-LIST			15	
da*	OFF-LIST	26			
deflector	OFF-LIST			10	
doin*	OFF-LIST	16			
dradis	OFF-LIST			18	
FBI	OFF-LIST		16	28	
fiancé	OFF-LIST	12			
frakkin	OFF-LIST			31	
frakking	OFF-LIST			16	
freaking	OFF-LIST	10			
ga*	OFF-LIST	14			
gettin*	OFF-LIST	11			
goanna*	OFF-LIST	18			
gosh	OFF-LIST	10			
hm*	OFF-LIST	11	16		
hmm*	OFF-LIST	39	45		
ho*	OFF-LIST	25	10		
hyperspace	OFF-LIST			11	
Illinois*	OFF-LIST		11		
intel	OFF-LIST		19		
intercept	OFF-LIST			10	
interphasic	OFF-LIST			11	
ipod	OFF-LIST	14			

Irene*	OFF-LIST	16			
Izzie*	OFF-LIST		51		
jeez	OFF-LIST	11			
kay*	OFF-LIST	15			
krypter	OFF-LIST			24	
malley	OFF-LIST		43		
mics	OFF-LIST				10
mmm*	OFF-LIST	10	38		
MRI	OFF-LIST		18		
Neill*	OFF-LIST			17	
nova	OFF-LIST			10	
nuke	OFF-LIST			14	
number:number*	OFF-LIST	33	30	13	
oi	OFF-LIST			21	
ow*	OFF-LIST	26	21		
pantyhose	OFF-LIST				14
Patterson*	OFF-LIST		11		
phaser	OFF-LIST			14	
Prudence*	OFF-LIST	20			
psychiatrist	OFF-LIST	28			
resonator	OFF-LIST			12	
ri*	OFF-LIST			20	
sensei	OFF-LIST	11			
shh*	OFF-LIST		25		
slicer	OFF-LIST	10			
subspace	OFF-LIST			40	
tardis	OFF-LIST			26	
teleport	OFF-LIST			16	
tumor	OFF-LIST		29		
txt*	OFF-LIST			58	
ULD	OFF-LIST			33	
ump	OFF-LIST	12			
wingman	OFF-LIST	11			
yo*	OFF-LIST		42		

Shaded areas are Teacher Talk Words.

*Interjections/inconsistencies and proper nouns have been taken out of the analysis because they are either known by the learners or do not contribute to this analysis.

Appendix B
Genre-Specific Learnable Words

Sitcoms	Dramas	Sci-fi programs	
Lame	Dynamite	Ambush	Wraith
Nap	Raft	Detonate	Disengage
Sibling	Ranch	Drone	Malfunction
Awesome	Spank	Infuse	Outpost
Porn	Surveillance	Hologram	Viper
Closet	Silo	Telepathy	Android
Balm	Vault	Traitor	Lingerie
Pumpkin	Freckle	Unicorn	Magma
Nickel	Fugitive	Weir	Nebula
VIP	Autopsy	Cairn	Raptor
Rickshaw	Directorate	Cellular	Shuttle
Bagel	Gunshot	Download	Adipose
Chiropractic	Syphilis	Flux	Roger
Salsa	Motel	Helm	Baseship(s)
Busboy	Shale	Interrogate	Cloaking
Jackass	Transceiver	Neural	Comm
Fiancé	Psych	Odyssey	Deflector
Freaking	Aneurism	Parchment	Dradis
Gosh	Bongo	Plasma	Frakkin
Ipod	Yin	Rendezvous	Frakking
Jeez	Cole	Visor	Hyperspace
Psychiatrist	Intel	Genocide	Intercept
Sensei	Malley	Hangar	Interphasic
Slicer	MRI	Lore	Krypter
Ump	Tumor	Teal	Nova
Wingman		Airlock	Nuke
		Artefact	Oi
		Conduit	Phaser
		DNA	Resonator
		Ensign	Subspace
		Fissure	Tardis
		Insurgent	Teleport
		Pituitary	ULD