Modern English Education, Vol. 14, No. 3, Fall 2013

Corpus-Based Vocabulary Analysis of *Friends*: Implications for Incidental Vocabulary Learning*

Shinwoong Lee
Hanyang University

Lee, Shinwoong. (2013). Corpus-based vocabulary analysis of *Friends*: Implications for incidental vocabulary learning. *Modern English Education, 14*(3), 1-21.

The primary purpose of the current study was bi-fold. It was to investigate what vocabulary size was necessary to comprehend *Friends*, one of the most popular American situation comedies, and to explore the number and percentage of potential words that might be incidentally learned through watching the program. The transcripts of two hundred twenty six episodes from *Friends* season one through ten were compiled and analyzed utilizing RANGE. It was found that the determined vocabulary sizes corresponding to 95% and 98% text coverage were 2,000 and 4,000 word families respectively, indicating that a relatively small number of word families were needed to comprehend *Friends* when compared to other TV programs. It was also revealed that *Friends* provided substantial opportunities for incidental vocabulary learning. Provided that learners knew about 2,000 word families and 95% coverage was enough for the comprehension of *Friends*, the number and proportion of the potential word families were 525 and 11.90% respectively. Meanwhile, assumed that learners had a vocabulary size of 4,000 word families and 98% coverage was adequate for comprehension of *Friends*, the number and percentage of the potential word families were 186 and 6.17 % respectively.

**[vocabulary size/TV programs/incidental vocabulary learning/
어휘크기/TV 프로그램/우연적 어휘학습**

* This was supported by the research fund of Hanyang University, 2013.

I. INTRODUCTION

A growing body of research (Freeman & Holden, 1986; Guariento & Morely, 2001; Sanderson, 1999) has indicated that using authentic materials can be conducive to teaching language and culture. Proponents of the use of authentic materials in language classrooms (Guariento & Morely, 2001; Little & Singleton, 1991; Peacock, 1997) argue that they can 1) increase learner motivation; 2) inform authentic target culture; and 3) provide exposure to the language that is naturally occurring in real life situations. Situated in EFL context where opportunities to interact with native speakers of English and to be exposed to real discourse are limited, language professionals have attempted to utilize TV programs and movies in their classrooms, and investigated the effect of their use on language teaching and learning (e.g., H. J. Chung & T. H. Lee, 2010; S. Y. Kim, 2010). Furthermore, as more and more of American TV dramas and movies are readily available on the Internet and cable TVs, they have become widely recognized as one of the major resources for learning English in Korea. In particular, a number of American dramas and situation comedies (e.g., *Friends*, *How I Met Your Mother*, and *Desperate Housewives*) have enjoyed popularity and their potential effectiveness on language teaching and learning have been explored extensively (e.g., J. H. Lee & J. W. Lee, 2012; J. Y. Seo & D. H. Lee, 2012).

However, despite their popularity and potentials in language teaching and learning, it has been relatively neglected to examine the lexical demand of TV programs and movies. Without knowing the vocabulary demand of the materials and expected pedagogical benefits, teachers may not be able to utilize them effectively in their classrooms. In this vein, the primary purpose of the current study was to examine the lexical demand of *Friends* that has been one of the most popular situation comedies and to investigate the number and percentage of potential words that the viewers of the program may incidentally learn though watching it.

II. LITERATURE REVIEW

1. Vocabulary Size for Comprehension

Determining the coverage of vocabulary needed to comprehend written and spoken discourse is crucial because it would inform L2 learners of the target vocabulary size they should reach. In this vein, a number of researchers suggested the text coverage that could lead to adequate comprehension of L2 texts (Hirsh & Nation, 1992; Hu & Nation, 2000; Laufer, 1989) and further researched what vocabulary size was necessary to reach the coverage suggested (Adolphs & Schmitt, 2003; Nation, 2006; Webb & Rodgers, 2009a,

2009b). Laufer (1989) found that 95% coverage was necessary for comprehension of a general academic L2 text. Hu and Nation (2000) examined the effect of the text coverage on reading comprehension of a fiction text and suggested that for largely unassisted reading, learners needed to know about 98% of running words of the text. Along the same lines, Hirsh and Nation (1992) and Nation (2001) suggested that 98% to 99% text coverage should be maintained for extensive reading. Analyzing CANCODE corpus, Adolphs and Schmitt (2003) found that the most frequent 3,000 word families (WFS, hereafter) accounted for about 96% of spoken discourse. Nation (2006) examined various written and spoken texts: five novels, the newspapers from five different corpora, graded readers, *Shrek* (a children's movie) and found that about 8,000 to 9,000 WFS were necessary for 98% text coverage of the written texts, while about 6,000 to 7,000 WFS were needed for 98% text coverage of the spoken texts. He suggested that 98% coverage was necessary for the comprehension of both spoken and written texts.

Particularly relevant to the vocabulary size necessary for comprehending British and American movies, Webb and Rodgers (2009b) analyzed 318 movies and found that most frequent 3,000 WFS plus proper nouns and exclamations provided 95 % coverage of both American and British movies. Meanwhile, to have 98% coverage, about 6,000 WFS were necessary for American movies and 7,000 WFS for British movies. In a similar vein, Webb and Rodgers (2009a) examined 88 American and British television programs and the findings echoed the result of their study of American and British movies: 3,000 WFS lead to about 95% coverage and 7,000 WFS accounted for about 98% of the running words of television programs. They argued that the combination of aural input and visual imagery would compensate for a relatively low text coverage (i.e., 95%) and suggested that if learners knew the most frequent 3,000 WFS and watched at least an hour per day, there was the potential for significant incidental vocabulary learning. There has been, however, only one study that directly investigated the effect of lexical coverage on listening comprehension in an experimental setting. Examining listening comprehension of four short passages with different lexical coverage, Bonk (2000) found a significant correlation between lexical coverage and listening comprehension and claimed that learners could have adequate listening comprehension even with 90% to 95% text coverage if listening comprehension strategies were effectively employed.

Taking all of the research findings into consideration, the vocabulary coverage for adequate comprehension of TV programs should range from 95% to 98%. Thus, in the current study, 95% and 98% were selected as the possible upper and lower boundaries and the determined vocabulary size corresponding to 95% and 98% coverage was interpreted as the vocabulary size needed to comprehend *Friends* and to have opportunities for incidental vocabulary learning through watching it.

2. Vocabulary Learning through Watching TV

A number of researchers studied incidental vocabulary learning through watching television. Pavakanun and d'Ydewalle (1992) investigated the effect of watching a cartoon on vocabulary learning. In a cartoon, they made the target words repeated at least five times and found that those who watched the cartoon in the target language without subtitles had better scores on a meaning recognition test than those who watched the movie with subtitles. They suggested that watching television could lead to substantial vocabulary learning if the repetition of target words was ensured. Similarly, in a study with Dutch children (8-12 years old), d'Ydewalle and Van de Poel (1999) examined the effect of watching a movie on incidental vocabulary learning with the target words being repeated at least four times in the movie. They found that vocabulary learning was much higher when presenting the target L2 in the soundtrack than presenting it in the subtitles.

With young L2 learners (seventh- and eighth-grade students), Neuman and Koskinen (1992) investigated the effect of three different activities (reading, watching television, and watching television with L2 subtitles) on vocabulary learning. They found that the learners who watched television with L2 subtitles had significantly higher scores on three measures of vocabulary than those who just read the transcripts. Also, the learners in the subtitled TV group had higher scores than those who just watched TV without subtitles. However, not all of the mean differences were statistically significant.

In a study with 246 Dutch children in grade four and six learning English, Koolstra and Beentjes (1999) examined the effect of watching television on L2 vocabulary acquisition. They found that the children who watched an English documentary with L1 (Dutch) subtitles outperformed those who just watched the program without subtitles and those who watched a Dutch television program (control group) on the vocabulary test. The result suggested that children could learn L2 vocabulary incidentally through watching subtitled television programs.

In sum, even though the effectiveness of the use of subtitles and how they should be presented are still inconclusive, most research findings indicated that watching television can be contributing to incidental vocabulary learning if multiple encounters of target words are ensured. Indeed, a number of studies (Elley, 1989; Rott, 1999; Saragi, Nation & Meister, 1978; Waring & Takaki, 2003; Webb, 2007) have recognized and confirmed the importance of repetition as a factor in incidental vocabulary learning. They suggested that six to twenty encounters may lead to vocabulary learning in written texts. In the current study, following their suggestion, the words that occurred more than ten times were considered to be the words that might be incidentally learned.

3. Statement of the Problem

As aformentioned, the use of movies and TV dramas has been encouraged in EFL classrooms as the use of authentic material has been cherished in L2 teaching and learning. However, only a few studies (M. S. Park & H. Jung, 2011; Webb & Rodgers, 2009a, 2009b) explore the lexical demand of movies and dramas. Determining the vocabulary size for understanding TV programs or movies is critical if they are intended to be used in classrooms. It is because knowing the lexical demand of the programs would help decide whether the vocabulary level of the programs is maintained at an optimal level at which effectiveness of incidental vocabulary learning can be maximized. It will also help learners identify the vocabulary size they should target by enabling them to recognize the gap between their vocabulary size and the vocabulary size needed for understanding the programs.

In addition, the number and percentage of potential words that can be learned through watching TV programs would also be worth investigating because it would help teachers and learners to set a vocabulary learning goal. However, to my knowledge, there has been only one study that investigated the number of encounters of the words in TV programs (i.e., Webb & Rodgers, 2009a). In this vein, the primary purpose of the current study was to determine the vocabulary size necessary for understanding *Friends* and compare the vocabulary size with that of other TV programs and movies, and also to investigate the number and percentage of potential words that might be learned incidentally through watching *Friends*. The following research questions guided the current study.

1) What vocabulary size is necessary for viewers to understand *Friends*?
2) Is there any difference between the vocabulary size needed for understanding *Friends* and other TV programs and movies?
3) What is the number and percentage of potential words that might be incidentally learned through watching *Friends*?

III. METHOD

1. Materials

The transcripts of two hundred twenty six episodes from *Friends* season one through ten were compiled to construct *Friends* Corpus. The transcripts of the show were taken from one of fan club websites, *Crazy for Friends* (http://www.liveinabox.com/*Friends*/). All of the transcripts were saved as text files for the purpose of analysis. The average number of

tokens per episode was 2,930 and the total number of tokens was 682,227 as shown in Table 1 below. The running time of each episode was 20 minutes and the total running time was 4520 minutes (75.3 hours). Twenty episodes were randomly selected and compared with the actual videos of the show and it was found that the transcripts were fairly accurate, even providing all the details of discourse including hesitations (e.g., *hmm*, *um*, *uh*), exclamations (e.g., *Ooh*, *Aaah*, *Dahaaa*), and repeats (*it's like*, *it's like*). They also provided the descriptions of the scenes and actors' performances (e.g., *The door buzzer sounds and Chandler gets it*.).

One of the primary reasons to choose *Friends* in constructing a corpus and analyzing its vocabulary was that it was not only one of the most popular situation comedies but also shared the core linguistic features that characterized natural conversation (Quaglio, 2009). Unlike other popular television dramas in which settings and vocabulary used were too area-specific (e.g., *CSI*, *Law and Order*, *ER*, etc.), the settings and types of interactions of *Friends* were relatively authentic to real-life situations, being a good resource for teaching and learning spoken English. Table 1 below shows the total number of episodes and tokens in each season along with the average number of tokens per episode.

TABLE 1
Composition of *Friends* Corpus

Season	Number of episodes	Number of token	Average number of token per episode
1 (1994-1995)	24	63,281	2,637
2 (1995-1996)	24	64,619	2,692
3 (1996-1997)	25	71,097	2,844
4 (1997-1998)	23	74,672	3,247
5 (1998-1999)	23	66,679	2,899
6 (1999-2000)	24	72,844	3,035
7 (2000-2001)	23	64,539	2,806
8 (2001-2002)	23	69,549	3,024
9 (2002-2003)	24	76,096	3,171
10 (2003-2004)	20	58,851	2,943
Total	226	682,227	2,930

2. Tool

The RANGE program was used to analyze the vocabulary of *Friends*. It was one of the most comprehensive vocabulary analysis tools, developed by Nation and Heatley (2002). It can provide the number and percentage of tokens, types, and word families in the fourteen 1,000-word-frequency lists that are used in the text. According to Nation and Heatley, the word lists from 1,000 to 14,000 were created on the basis of the frequency word lists of British National Corpus (BNC) and three factors were taken into consideration when selecting the words: range, frequency, and dispersion. First, in order to check a range of words, they checked whether they were used in diverse corpora (e.g., LOB, FLOB, Brown, Frown, Kohlapur, Wellington Written, Wellington Spoken and LUND corpora). Then, they measured the frequency of words and examined how evenly the words were dispersed in each corpus. The word families in the lists were created following Bauer and Nation's (1993) classification of word families; the word families included stems, and inflections and about 80 derivational affixes. The RANGE program and the fourteen word lists of BNC can be downloaded from Paul Nation's Web site: http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx.

3. Procedures and Analysis

Before analyzing the vocabulary of *Friends* using RANGE, all of the contracted forms (e.g., *wanna*, *outta*), apostrophes (e.g., *it's*, *sayin'*) and hyphenated words (e.g., *cross-eyed*) were changed to conform to the spellings used in the word lists of RANGE. For example, *let's*, *it's been*, *gotta*, *gonna*, *kinda*, and *'cause* were changed to *let us*, *it has been*, *got to*, *going to*, *kind of* and *because*. In addition, the words of descriptions of scenes, actor's performance, and speakers' names were also removed from the transcripts (e.g., *entering from her bedroom carrying a present*, *Scene: Monica and Rachel's, Carol and Susan are showing off Ben to the gang*.) for a more accurate vocabulary analysis.

Then, the revised transcripts of *Friends* were analyzed drawing on RANGE. The result of analysis displayed the number of type, token, and word family used in each word band (from 1,000 through 14,000) and their percentage out of the total number of each word counting unit (i.e., type, token, and word family). The word band 15 and 16 were the list of proper nouns and exclamations respectively and it was assumed that they could be understood easily with a minimal learning burden.

The words that were not in the fourteen 1,000-word-family list of BNC were manually analyzed and then regrouped. First, they were searched on two online dictionaries (at www.daum.net and www.naver.com), and when they were found to be actual words, they were categorized as "not in the list." Then, the rest of the words were re-examined to see whether they could be potentially subsumed under the word bands. For example, even

though *opposable* was not in the list, it was subsumed under the 1,000 word band which *oppose* belonged to, assuming that the meaning of *opposable* could be easily guessed once they knew the meaning of *oppose*.

As for compound words (e.g., *showtime*, *matchmaker*), they were analyzed according to the roots that comprised them. For example, in the case of *windshield*, *wind* and *shield* were searched on the word lists separately and the word band that each word belonged to was identified separately (*wind*: 1,000 word band, *shield*: 3,000 word band) and then the higher word band between two was chosen, subsuming *windshield* under the 3,000 word band. The rest of words were searched on *Wikipedia*, the most comprehensive online encyclopedia available, to see whether they were proper nouns such as the names of people and places. If the words were found to be proper nouns, they were re-grouped into "proper nouns." Meanwhile, all of marginal words such as *ahh*, *ahaha*, and *oh* were categorized as "exclamation." Finally, the words that were found neither in the dictionaries nor in the encyclopedia were excluded since they made up only a small proportion of the text (0.07%, 490 tokens) and thus were negligible.

Consequently, a certain number of words were added to each word band and the token of those words were also added up accordingly when the text coverage of each word band was calculated. In addition, the number of encounters of word families at each word band (from 3,000 to 14,000 word level) was calculated. It was to measure how often words were repeated in *Friends* and thus to know the number and percentage of potential words that could be incidentally learned through watching the program.

IV. FINDINGS AND DISCUSSION

1. Vocabulary Size for Comprehending *Friends*

As shown in Table 2 below, it was found that 682,222 tokens, 13,795 types, and 8,375 WFS were used in *Friends*. The words were spread over fourteen most frequent 1,000 WFS of the BNC list and beyond. As the word frequency decreased, the coverage of word band decreased sharply, and by the 4th 1,000 level, the coverage became lower than 1%, indicating a relatively low significance of knowing 4th 1,000 WFS onwards. The first 1,000 WFS accounted for 88% of the text, covering 600,334 running words and the second 1,000 accounted for about 3%, covering 20,393 running words.

Table 3 below summarized the text coverage of each word band, the coverage including proper nouns and exclamations and the coverage without them. Supposed that proper nouns and exclamations could be easily understood with a minimal learning burden, 95% coverage could be reached with 2,000 WFS and 98% coverage with 4,000 WFS. The result

also indicated that learners who knew the most frequent 2,000 WFS to 4,000 WFS might have an opportunity for incidental vocabulary learning while watching *Friends*.

The number of proper nouns and exclamations made up a relatively large proportion of the total number of running words. The coverage of proper nouns was 2.52%, accounting for 171,831 tokens and that of exclamation was 2.76%, covering 18,833 tokens. The coverage of proper nouns and exclamations were the fourth and the third highest respectively following the first and second most frequent 1,000 WFS. As shown in Table 3, without knowing them, the minimum text coverage (95%) for comprehending *Friends* could not be reached even with 14,000 WFS, which clearly showed the importance of knowing proper nouns and exclamations.

TABLE 2

Token, Type, and Word Family at Each Word Level in *Friends*

Word band	Token (%)	Type (%)	Family (%)
1st 1000	600,334 (88.00)	3,136(22.73)	989(11.81)
2nd 1000	20,398 (2.99)	2,289(16.59)	913(10.90)
3rd 1000	9,078 (1.33)	1,221(8.85)	807(9.64)
4th 1,000	5,269 (0.77)	1,047(7.59)	642(7.67)
5th 1,000	3,048 (0.45)	761(5.52)	523(6.24)
6th 1,000	2,343 (0.34)	563(4.08)	409(4.88)
7th 1,000	1,162 (0.17)	397(2.88)	320(3.82)
8th 1,000	762 (0.11)	330(2.39)	267(3.19)
9th 1,000	643 (0.09)	283(2.05)	232(2.77)
10th 1,000	679 (0.10)	225(1.63)	185(2.21)
11th 1,000	412 (0.06)	188(1.36)	153(1.83)
12th 1,000	266 (0.04)	131(0.95)	109(1.30)
13th 1,000	211 (0.03)	110(0.80)	94(1.12)
14th 1,000	173 (0.03)	82(0.59)	71(0.85)
Proper Noun	17,183 (2.52)	1,909(13.84)	1,883(22.48)
Exclamation	18,833 (2.76)	414(3.00)	127(1.52)
Not in the list	1433 (0.21)	709(5.14)	651(7.77)
Total	682,227	13,795	8375

TABLE 3

Text Coverage Figures for *Friends*

Word band	Token	Coverage (%)	Cumulative coverage without proper noun & exclamation (%)	Cumulative coverage with proper noun & exclamation (%)
1st 1000	600334	88.00	88.00	93.28
2nd 1000	20398	2.99	90.99	96.27*
3rd 1000	9078	1.33	92.32	97.60
4th 1,000	5269	0.77	93.09	98.37**
5th 1,000	3048	0.45	93.54	98.82
6th 1,000	2343	0.34	93.88	99.16
7th 1,000	1162	0.17	94.05	99.33
8th 1,000	762	0.11	94.16	99.44
9th 1,000	643	0.09	94.26	99.53
10th 1,000	679	0.10	94.36	99.63
11th 1,000	412	0.06	94.42	99.69
12th 1,000	266	0.04	94.45	99.73
13th 1,000	211	0.03	94.49	99.76
14th 1,000	173	0.03	94.51	99.79
Proper Noun	17183	2.52		
Exclamation	18833	2.76		
Not in the list	1433	0.21		
Total	682,227			

* 95% coverage cut-off point, **98% coverage cut-off point

2. Text Coverage of *Friends* and Other TV Programs and Movies

Table 4 below displayed the comparison of text coverage between *Friends* and other TV programs and movies examined in Webb and Rodgers' (2009a, 2009b) study. The coverage of 1st 1,000 WFS in *Friends* (88.00%) was higher than the ones that were found in 88 television programs (85.11%, Webb & Rodgers, 2009a) and in 318 movies (86.52%, Webb & Rodgers, 2009b). The coverage of 3,000 WFS in *Friends* (92.32%) was higher than the one found in 88 TV programs (91.46%) and similar to the one found in 318 movies (92.39%).

As shown in Table 4 below, when assumed proper nouns and exclamations could be understood easily, 2,000 WFS provided more than 95% (i.e., 96.27%) coverage of the running words in *Friends* and with 4,000 WFS, 98% of the running words could be known. On the other hand, a higher level of vocabulary size was necessary to understand American and British movies and TV programs according to the findings of Webb and Rodgers' (2009a, 2009b) study. They reported that the most frequent 3,000 WFS provided about

95% coverage of the running words of 88 TV programs and 318 movies, while 7,000 and 6,000 WFS were needed to have 98% coverage of the running words respectively. It appeared that lexical demand for watching *Friends* was lower than that of other American and British TV programs. It was primarily because the coverage of high-frequency words (i.e, most frequent 3,000 WFS) in *Friends* was higher than that in other TV programs, and exclamations made up a relatively larger proportion of the running words in *Friends*, when compared to those of other TV programs (See Table 4 below, 2.76% vs. 1.03% and 0.70%). They all contributed to lowering the vocabulary size that can lead to 95% and 98% text coverage.

TABLE 4

Text Coverage Figures for *Friends*, TV Programs, and Movies

Word band	*Friends*		88 TV programs (Webb & Rodgers, 2009a)		318 movies (Webb & Rodgers, 2009b)	
	A[1]	B[2]	A	B	A	B
1st 1000	88.00	93.28	85.11	89.10	86.52	89.89
2nd 1000	90.99	96.27*	89.53	93.52	90.67	94.04
3rd 1000	92.32	97.60	91.46	95.45*	92.39	95.76*
4th 1,000	93.09	98.37**	92.76	96.75	93.67	97.04
5th 1,000	93.54	98.82	93.47	97.46	94.33	97.70
6th 1,000	93.88	99.16	93.96	97.95	94.78	98.15**
7th 1,000	94.05	99.33	94.28	98.27**	95.03*	98.40
8th 1,000	94.16	99.44	94.49	98.48	95.22	98.59
9th 1,000	94.26	99.53	94.69	98.68	95.38	98.75
10th 1,000	94.36	99.63	94.84	98.83	95.51	98.88
11th 1,000	94.42	99.69	94.97	98.96	95.61	98.98
12th 1,000	94.45	99.73	95.06*	99.05	95.68	99.05
13th 1,000	94.49	99.76	95.13	99.12	95.75	99.12
14th 1,000	94.51	99.79	95.17	99.16	95.79	99.16
Proper Noun	2.52		2.96		2.67	
Exclamation	2.76		1.03		0.70	

* 95% coverage cut-off point, **98% coverage cut-off point

[1] A refers to the coverage without proper nouns and exclamations.

[2] B refers to the coverage including proper nouns and exclamations.

3. Number and Percentage of Encounters

Table 5 below displayed the number and percentage of encounters out of the total number of word families at each word band. As shown in the table, the number and percentage of word families that were encountered once or twice tended to increase as the word level increased. For example, the percentage of word families that were encountered once at 3,000 word level was 16.93 and the percentage at 4,000 word level was 26.95, showing a significant increase. In contrast, the number and percentage of word families that are repeated more than five times tended to decrease as the word level increased and this tendency could be clearly observable in the word families that were encountered more than eight times. For example, the percentage of word families that were encountered more than ten times at 3,000 word level was 28.75 and the percentage at 4,000 word level was 16.67, displaying a substantial decrease. In sum, the number of low-encountered word families tended to increases and the number of high-encountered word families tended to decrease as the word level increased. The result suggested that *Friends* offered a strong potential for large incidental vocabulary gains especially for the learners who had a vocabulary size of 2,000 WFS and targeted the words at 3,000 word level since the proportion of the word families that were encountered more than ten times was the largest at this word level (28.75%).

TABLE 5

Number and Percentage of Encounters out of WFS at Each Word Band

	1 (%)	2 (%)	3-4 (%)	5-7 (%)	8-9 (%)	10+ (%)	Total
3rd 1,000	137(16.98)	106(13.14)	123(15.24)	149(18.46)	54(6.69)	232(28.75)	807
4th 1,000	173(26.95)	101(15.73)	131(20.40)	97(15.11)	33(5.14)	107(16.67)	642
5th 1,000	181(34.61)	99(18.93)	89(17.02)	71(13.58)	22(4.21)	61(11.66)	523
6th 1,000	168(41.08)	73(17.85)	67(16.38)	46(11.25)	7(1.71)	48(11.74)	409
7th 1,000	146(45.63)	77(24.06)	51(15.94)	20(6.25)	6(1.88)	20(6.25)	320
8th 1,000	143(53.56)	53(19.85)	37(13.68)	14(5.24)	6(2.25)	14(5.24)	267
9th 1,000	135(58.19)	42(18.10)	26(11.21)	15(6.47)	5(2.16)	9(3.88)	232
10th 1,000	95(56.21)	45(24.32)	21(11.35)	13(7.03)	2(1.08)	9(4.86)	185
11th 1,000	86(53.21)	29(18.95)	23(15.03)	6(3.92)	3(1.96)	6(3.92)	153
12th 1,000	58(53.21)	22(20.18)	19(17.43)	6(5.50)	2(1.83)	2(1.83)	109
13th 1,000	56(59.57)	17(18.09)	15(15.96)	2(2.13)	0(0.00)	4(4.26)	94
14th 1,000	34(47.89)	19(26.76)	10(14.08)	6(8.45)	0(0.00)	2(2.82)	71

Table 6 below showed the number of encounters with 3,000 to 14,000 word-level families and the percentage of total number of encounters according to the number of encountering (one through ten) in *Friends*. The reason for choosing these particular word bands was that 2,000 WFS were found to provide 95% coverage in *Friends* and thus by observing the number of encounters in these word bands, the number of potential words that the learners who knew 2,000 WFS might incidentally learn could be measured. As shown in Table 6 below, the words that occurred once or twice made up 58.68% of the total number of word families used from 3,000 to 14,000 word levels and the percentage of word families that occurred three to seven times was 26.01%. When assumed that leaners had a lexical knowledge of 2,000 WFS and 95% coverage would be enough for the comprehension of *Friends*, learners might have opportunities to learn a relatively large number of unknown words. The proportion of the words that occurred more than ten times was 11.90% (525 WFS), which would be the percentage of the potential words that might be learned incidentally. If the word families that occurred eight to nine times were included, the number and percentage of the potential words would increase significantly (15.31%, 677 WFS).

TABLE 6
Number of Encounters with 3,000-14,000-level Word Families

	1	2	3-4	5-7	8-9	10+
3rd 1,000	137	106	123	149	54	232
4th 1,000	173	101	131	97	33	107
5th 1,000	181	99	89	71	22	61
6th 1,000	168	73	67	46	7	48
7th 1,000	146	77	51	20	6	20
8th 1,000	143	53	37	14	6	14
9th 1,000	135	42	26	15	5	9
10th 1,000	95	45	21	13	2	9
11th 1,000	86	29	23	6	3	6
12th 1,000	58	22	19	6	2	2
13th 1,000	56	17	15	2	0	4
14th 1,000	34	19	10	6	0	2
Not in the list	413	111	71	33	12	11
Number of total WFS	1825	794	683	478	152	525
Percentage (%)	40.89	17.79	15.30	10.71	3.41	11.90

Table 7 below displayed the number of encounters with 5,000 to 14,000 word-level families and the percentage of total number of encounters according to the number of encountering (one through ten) in *Friends*. Similar to the reason explained above, the purpose of choosing 5,000 to 14,000 word bands was that 4,000 WFS were found to provide 98% coverage of *Friends* and thus by calculating the number of encounters in these word bands, the number of potential words that the learners who had a vocabulary size of 4,000 WFS might incidentally learn could be measured. As shown in Table 7 below, the words that occurred once or twice made up 69.75% of the total number of word families from 5,000 to 14,000 word levels and the percentage of the words that occurred three to seven times was 21.93%. When supposed that 98% was the threshold coverage for adequate comprehension of *Friends* and incidental vocabulary learning, the learners who had a vocabulary size of 4,000 WFS might have an opportunity to learn 6.17 % (186 WFS) of the word families from 5,000 to 14,000 word levels. If the words that occurred eight to nine times are included, the number and percentage of the potential words for incidental learning would increase moderately (8.33%, 251 WFS).

However, it should be noted that repeated exposure to words does not necessarily bring about vocabulary learning. Rather, it would raise learners' consciousness and lead them to pay more attention to the particular words, increasing the likelihood of learning those words in part. Thus the result of the analysis including numbers and percentages provided in the current study should be interpreted with caution.

TABLE 7

Number of Encounters with 5,000-14,000-level Word Families

	1	2	3-4	5-7	8-9	10+
5th 1,000	181	99	89	71	22	61
6th 1,000	168	73	67	46	7	48
7th 1,000	146	77	51	20	6	20
8th 1,000	143	53	37	14	6	14
9th 1,000	135	42	26	15	5	9
10th 1,000	95	45	21	13	2	9
11th 1,000	86	29	23	6	3	6
12th 1,000	58	22	19	6	2	2
13th 1,000	56	17	15	2	0	4
14th 1,000	34	19	10	6	0	2
Not in the list	413	111	71	33	12	11

Number of total WFS	1515	587	429	232	65	186
Percentage (%)	50.27	19.48	14.23	7.70	2.16	6.17

4. Discussion and Implications

It was found that the minimum vocabulary size to understand *Friends* was 2,000 WFS, covering 95% of the running words, and at least 4,000 WFS were necessary to have 98% coverage, which could possibly lead to adequate comprehension of *Friends*. As reported in the previous section, the lexical demand of *Friends* was lower than that of other TV programs and movies, and this implied that *Friends* could be used to the learners with a relatively low vocabulary size. As Webb and Rodgers (2009a) argued, if 95% text coverage would be enough to understand TV programs with visual input being provided, *Friends* could be used for the learners who knew about 2,000 WFS. One of the problems of using other TV programs and movies was that the vocabulary size necessary to understand them was much higher than learners' vocabulary size. As found in previous studies (Nation, 2006; Webb & Rodgers, 2009a, 2009b), ideally 6,000 and 7,000 WFS were necessary to understand TV programs and movies. However, many EFL learners may not reach that vocabulary level yet, and thereby those TV programs and movies could not be utilized in their classrooms. In this sense, *Friends* provided better learner accessibility and learnability since its lexical demand was lower than that of other TV programs and movies.

It was also found that the learners who had a vocabulary size of 2,000 WFS would have a chance to learn 11.90% (525 WFS) of the total word families used at 3,000 to 14,000 word levels when assumed that 95% coverage would be sufficient for comprehending *Friends*. In addition, when supposed that 98% was the threshold coverage for adequate comprehension of *Friends* and incidental vocabulary learning, the learners who had a vocabulary size of 4,000 WFS may have an opportunity to learn 6.17 % (186 WFS) of the total number of word families used at 5,000 to 14,000 word levels. It appeared that *Friends* provided more opportunities for incidental vocabulary learning than other TV programs. In fact, the percentage of word families occurring more than ten times in *Friends* was higher than in other TV programs examined in Webb and Rodgers' (2009a) study. They reported that about 4% of the word families from 5,000 to 14,000 word bands occurred more than ten times in all of the TV programs investigated. In particular, only 1% of the word families were encountered more than ten times in the genre of dramas and situation comedies. The result of data analysis in the current study clearly suggested that *Friends* would provide substantial opportunities for incidental vocabulary learning.

Based on the findings of the current study, a few implications can be provided for language professionals in relation to the use of TV programs and movies in their classrooms. First, it is crucial to know what vocabulary size is necessary to understand them because it will partially determine the degree of difficulty of materials and play an important role in deciding whether they can be used in the classroom or not. In other words, if the vocabulary level that is required to comprehend them is way beyond learners' level, they may not be justifiably used in the classroom. For this reason, appropriateness of the materials with regards to their lexical demand should be checked in advance. Also, examining how often words are repeated in TV programs and movies can be conducive to setting a vocabulary learning goal, assisting vocabulary instruction in general. Possibly, the list of target words can be previewed explicitly so that learners can focus on the words, which may increase the probability of learning those words.

In addition, as found in the current study, the percentage of proper nouns and exclamations was higher than that found in the previous studies (Webb & Rodgers, 2009a, 2009b). It evidently showed the importance of recognizing proper nouns and marginal words such as exclamations. Even though proper nouns including names of people and places are often assumed that they can be understood easily, they can indeed impede listening comprehension processes. Taking characteristics of spoken discourse into consideration, one of which is that the stream of speech will just continue to flow unlike reading in which one can just stop and go back to reread (Brown, 2001), the failure to recognize proper nouns immediately would hinder learners from processing further information. Therefore, it is essential to introduce major characters and places and have learners be familiar with them in advance to promote listening comprehension of TV programs and movies.

Finally, as Quagglio (2009) argued, *Friends* can be a valuable asset to an effort to bring natural conversation to the classroom and it is more so when considering the lack of readily available spoken corpora and the difficulty in collecting spoken data. As found in the current study, the lexical demand of *Friends* was lower than that of other TV programs and movies, and this implies that more English learners including the learners who have a relatively low vocabulary size might be able to use *Friends* as a reference to spoken English. They could not only utilize authentic examples of spoken English, but also learn numerous aspects of American cultures that are interwoven in *Friends*.

V. CONCLUSION

It was found that the determined vocabulary sizes corresponding to 95% and 98% coverage were 2,000 and 4,000 WFS respectively, indicating that a relatively small number

of word families were needed to comprehend *Friends* when compared to the other TV programs. It was also revealed that when assumed that leaners had a lexical knowledge of 2,000 WFS and 95% coverage would be enough for the comprehension of *Friends*, the proportion of the potential words that might be incidentally learned through watching it was 11.90% (525 WFS). Meanwhile, when presupposed that learners had a vocabulary size of 4,000 WFS and 98% was the coverage for adequate comprehension of *Friends*, the percentage of the potential words was 6.17 % (186 WFS). Indeed, the result of analysis showed that *Friends* offered better leaner accessibility and learnability since its lexical demand was lower than other TV programs and provided substantial opportunities for incidental vocabulary learning as well. Even though vocabulary learning through reading has been strongly advocated by a number of researchers (e.g., Laufer, 1989; Nation, 2001, 2006), it should also be admitted that people tend to spend more time watching television than reading books. Furthermore, a number of American TV programs including dramas and situation comedies have become increasingly popular among Korean learners of English. This suggests that they could be utilized for learning vocabulary in and out of classrooms. This being the case, *Friends* has a strong potential to be used for learning vocabulary while providing examples of authentic uses of spoken English.

As noted earlier, the results should be interpreted with caution. Even though the findings of studies (Elley, 1989; Rott, 1999; Saragi, Nation & Meister, 1978; Waring & Takaki, 2003; Webb, 2007) on vocabulary learning suggested that six to twenty encounters may lead to vocabulary learning, exposing to unknown words repeatedly in context might just let learners raise their consciousness and lead them to pay more attention to the words. Vocabulary learning is an accumulative and time-consuming process in which learners typically acquires receptive lexical knowledge at the beginning, trying to recognize written and spoken forms and match them to the corresponding meanings and then move on to more productive lexical knowledge (e.g., grammatical patterns and collocations) later. For this reason, it is premature to say that encountering unknown words in context more than ten times would lead to incidental vocabulary learning. In order to identify the true effect of repetition of words on vocabulary learning, an experimental study should be conducted in which the unique effect of repetitions on vocabulary learning can be examined by controlling other nuisance factors.

In addition, it should also be noted that even though lexical coverage is one of the most important factors that affect listening comprehension (Bonk, 2000; Kelly, 1991; Mecartty, 2000), 95% or 98% coverage does not necessarily lead to 95% or 98% comprehension. It is because there are many other factors that can considerably influence the comprehension of spoken texts such as prior knowledge (Bacon, 1992; Dunkel & Chang, 1992) and speech rate (Brown & Yule, 1983; Conrad, 1989; Schmidt-Rinehart, 1994).

Another problem arises from the assumption that if learners know the meaning of separate word of word chunks, they know the meaning of the whole sequences. In fact, a large number of phrasal verbs and idiomatic expressions are used in TV programs including *Friends*, and when calculating the coverage, it is often assumed that if we know separate words that compose phrasal verbs and idiomatic expressions, we know the meaning of the whole sequences. However, it is very questionable because they are typically non-transparent in meaning. For example, knowing the meaning of *put* and *off* respectively does not necessarily mean knowing the meaning of *put off*. Similarly, knowing the meaning of *with*, *a*, *grain*, *of*, and *salt* separately does not imply knowing the meaning of the whole sequence, *with a grain of salt*. Thus, 95% or 98% text coverage should be interpreted with caution.

REFERENCES

Adolph, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, *24*(4), 425-438.

Bacon, S. (1992). The relationship between gender, comprehension, processing strategies, and cognitive and affective response in foreign language listening. *The Modern Language Journal, 76*(2), 160-178.

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography, 6*(4), 253-279.

Brown, D. (2001). *Teaching by principles: An interactive approach to language pedagogy*. White Plains, NY: Addison Wesley Longman.

Brown, G., & Yule, G. (1983). *Teaching the spoken language*: *An approach based on the analysis of conversational English*. Cambridge: Cambridge University Press.

Bonk, W. (2000). Second language lexical knowledge and listening comprehension. *International Journal of Listening*, *14*(1), 14-31.

Chung, Hye-Jin, & Lee, Tae-Hyung. (2010). Improving listening and speaking skills using different types of media and techniques through movies. *STEM Journal, 11*(1), 149-171.

Conrad, L. (1989). The effect of time-compressed speech on native and ESL listening comprehension. *Studies in Second Language Acquisition, 11*(1), 1-16.

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

Dunkel, P., & Chiang, C. (1992). The effect of speech modification, prior knowledge, and listening proficiency on EFL learning. *TESOL Quarterly, 26*(2), 345-374.

Elley, W. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 24*(2), 174-187.

Freeman, D., & Holden, S. (1986). Authentic listening materials. In S. Holden (Ed.), *Techniques of teaching* (pp. 67-69). London: Modern English Publication.

Guariento, W., & Morley, J. (2001). Text and task authenticity in the EFL classroom. *ELT Journal, 55*(4), 347-353.

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.

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

Kelly, P. (1991). Lexical ignorance: The main obstacle to listening comprehension with advanced foreign language learners. *IRAL, 29*(2), 135-149.

Kim, So-Yeon. (2010). A study of writing through verb phrase awareness and movies. *STEM Journal, 11*(2), 45-66.

Koolstra, C., & Beentjes, 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.

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

Lee, Ji-Hyun, & Lee, Ja-Won. (2012). A study of language learning using depth of processing: Based on the character Ross in *Friends*. *STEM Journal, 13*(1), 129-148.

Little, D., & Singleton, D. (1991). Authentic texts, pedagogical grammar and language awareness in foreign language learning. In C. James & P. Garret (Eds.), *Language awareness in the classroom* (pp. 123-132). London: Longman.

Mecartty, F. (2000). Lexical and grammatical knowledge in reading and listening comprehension by foreign language learners of Spanish. *Applied Language Learning, 11*(2), 323-348.

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

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

Nation, P., & Heatley, A. (2002). Range: A program for the analysis of vocabulary in texts[software]: Downloadable from the World Wide Web: http://www.vuw.ac.na/lals/staff/paul-nation/nation.aspx

Neuman, B., & Koskinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading Research Quarterly, 27*(1), 95-106.

Park, Myong-Su, & Jung, Hyun. (2011). Corpus-based vocabulary analysis of American TV drama series. *STEM Journal, 12*(2), 75-94.

Pavakanun, U., & d'Ydewalle, G. (1992). Watching foreign television programs and language learning. In F. Engel, D. Bouwhuis, T. Bösser & G. D'Ydewalle (Eds.), *Cognitive modeling and interactive environments in language learning* (pp. 193-198). Berlin: Springer.

Peacock, M. (1997). The effect of authentic materials on the motivation of EFL learners. *ELT Journal, 51*(2), 144-156.

Quaglio, P. (2009). *Television dialogue: The sitcom Friends vs. natural conversation*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition through reading. *Studies in Second Language Acquisition, 21*(4), 589-619.

Sanderson, P. (1999). *Using newspapers in the classroom*. Cambridge: Cambridge University Press.

Saragi, T., Nation, P., & Meister, G. (1978). Vocabulary learning and reading. *System, 6*(2), 72-78.

Schmidt-Rinehart, B. (1994). The effects of topic familiarity on second language listening comprehension. *The Modern Language Journal, 78*(2), 179-189.

Seo, Ji-Young, & Lee, Dong-Han. (2012). The role of language play in awareness-raising tasks: Using captions from Desperate Housewives. *STEM Journal, 13*(1), 57-80.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*(1), 1-27. Retrieved from March 3, 2013, from the World Wide Web: http://nflrc.hawaii.edu./Rfl/October2003/waring/waring.html.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28* (1), 46-65.

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

Webb, S., & Rodgers, M. (2009b). The lexical coverage of movies. *Applied Linguistics, 30*(3), 407-427.

Examples in: English
Applicable Languages: English
Applicable levels: Secondary/ Tertiary

Shinwoong Lee
Department of English Language and Literature
Hanyang University
Sungdong-ku Hangdang-dong 17
Seoul, 133-791
Tel: (02) 2220-0745/H.P.: 010-2259-9393
Email: shinwoonglee@hanyang.ac.kr

Received 15 June 2013
Revised 31 July 2013
Accepted 23 August 2013