TV DRAMA AS AN EFFECTIVE METHOD FOR LEXICAL INSTRUCTION: A CORPUS-BASED LEXICAL RICHNESS STUDY*

Sibo Chen
University of Victoria

An effective lexical instruction method serves as an important component in any English as an Additional Language (EAL) program. In recent years, the advancement of Internet technologies has dramatically changed the online English video distribution. Now many TV dramas in the United States can be watched by Chinese EAL learners with the support of bilingual (English/Mandarin) subtitling. Thus it is possible to used English TV Drama as an effective method for EAL lexical instruction. This article reports a corpus-based study in which the popular US TV sitcom *Friends* was focused. The research question was whether the language in the sitcom can provide sufficient lexical input for Chinese EAL learners. With the utilization of corpus-based methodology, the study found that most tokens in the scripts of the sitcom were in concordance with the most frequent lexical items in the British National Corpus (BNC) and the Academic Word List (AWL), which indicated that English TV sitcoms may serve as an effective tool for EAL lexical instructions.

1. Introduction

With the enhancement of technology, especially the development of the Internet in recent years, technology is playing a vital role in Foreign Language Education. As suggested by Blake (2008), the language classroom in the 21st century is developing towards a digital version. Stimulated by the Internet technology, online videos in English are widely watched by learners of English as an Additional Language (EAL). In the past decade, TV dramas from the United States have received increasing attention among EAL learners in China. One single research of "American TV Drama" in the Chinese search engine Baidu can generate more than 27 million results.

Although several previous studies have suggested the potential benefit of audiovisual TV programs for SLA, such as Bahrani (2011), Bird (2005), Inglese, Mayer, and Rigotti (2007), and William and Thorne (2000), English TV dramas have rarely been studied from a lexical instruction perspective, which is disappointing and thus forms a worthwhile research topic.

The focus of the current study was the U. S. TV sitcom *Friends*. As one of the earliest U. S. TV sitcoms imported into China in the late 1990s, *Friends* is generally regarded as the most popular English TV show in China. Its popularity and vivid play scripts have attracted many Chinese applied linguists. Previous studies concerning *Friends* in language education have been done from many perspectives including collocation and idioms (Ye, 2005), hedges (Hu, 2007),

affective metaphor (Chen, 2011), and so on. However, one crucial aspect, the lexical richness of this TV sitcom has not yet been studied. Based on such research gap, the current study incorporated the corpus linguistics method and explored whether the sitcom *Friends* can provide sufficient recourse for EAL learners' lexical development.

2. Literature Review
2.1 Lexical Instruction and the Lexical Approach

Learners' knowledge of a foreign language's lexicons has a fundamental influence on their performance in that language. Traditional teaching methods (e.g. the Grammar-Translation Approach and the Audio-Lingual Method) apply the drill-practice method for lexicon memorization mechanically (Cook, 2000), which is reported as detrimental for learners' initiative as well as enthusiasm in language learning (Littlewood, 1982; Rivers, 1981; Celce-Murcia, 2001). In recent years, vocabulary instruction has been relegated to a secondary position with the widespread application of Communicative Language Teaching (CLT), and learners are expected to expand their lexical inventory along with the communicative interactions using the target language (Krashen & Terrell 1983). However, several studies, such as Gass (1988) and Nation (1993), show opposite results. According to these studies, learners' lexicon inventory is somewhat limited despite being involved in CLT programs for a certain period. As discussed in Nation (1993), an inventory of 3000 words is a crucial threshold in a learner's second language acquisition process. After this threshold, learners will gradually focus more on the content of the foreign language, which can be efficiently facilitated by the CLT method. Thus, how to facilitate L2 learners to achieve the 3000-word inventory should be a central consideration in lexical pedagogy designs.

In the past two decades, one influential method in teaching second language vocabulary is Lewis' Lexical Approach (Lewis, 1993, 1997). The basic concept of the Lexical Approach is that vocabulary is prioritized to grammar *per se*. Lewis argues that learning a language consists of being able to comprehend and produce the lexical phrases in that language. Thus, if students were taught to perceive lexical chunks in that language, they would be able to understand the language patterns (grammar) of the target language and use the target language meaningfully. Instructors in the Lexical Approach are supposed to concentrate on fixed expressions that occur frequently in the target language's daily conversations (Lewis 1993).

The Lexical Approach has received contradictory reviews since its appearance. Hall (1994), in a review of Lewis' work, emphasizes the approach's ambitious perspective and its innovational ideas. By contrast, the approach has also been criticized for its lack of theoretical ground and inadequacy of teaching personality structures (e.g. Block 1995; Westen 1996). As a response to the questions proposed by other scholars, Lewis published *Implementing the Lexical Approach: Putting Theory into Practice* in 1997, in which he exemplifies how the conceptions of the Lexical Approach can be effectively applied in real language classrooms. Since then, the Lexical Approach has spread widely among TESL/TEFL professionals. One recent development, for instance, is the advancement of collocation studies (Lewis 2000). To sum up, the Lexical Approach makes a significant contribution to our understanding of the status of lexicons in the second language development procedure.

2.2 Corpus-based Vocabulary Research and Previous Studies in Media and Language Instruction

Corpus linguistics is becoming a prevalent field in SLA research along with the advancement of corpus construction and corpus analysis software (Bennett, 2010; Reppen, 2010; Huang, 2008). With the implementation of large quantity of text data, corpus linguistics is able to reveal frequency and collocation patterns of target texts and thus can be a useful directory for language learners. For instance, when an Asian learner is learning English as an Additional Language (EAL), with so many electronic corpora available online, he/she can get a close look at a particular word, to compare different texts, and to learn frequently used phrases in academic articles with the assistance of well-organized corpora such as the British National Corpus (BNC), the Bank of English, or the American National Corpus (ANC) (Reppen, 2010). Another notion that has been developed in recent years is Data Driven Learning (DLL) (Gavioli & Aston, 2001; Johns, 1997), in which students act as "language detectives" to actively participate in the discovery of language patterns.

In respect to corpus-based lexical instruction, one key figure is Paul Nation who actively promotes the development of the Academic Word List (Coxhead, 2000) and its relevant applications in vocabulary instructions for both EAL reading and speaking (see Laufer & Nation, 1995; Nation, 1993; Nation & Coxhead, 2001 for more information). One key concept in numerous studies by Nation is lexical richness, which describes the range of words in a target text. In the current study, the lexical richness of the scripts of the TV sitcom *Friends* was investigated with the application of the corpus software tool developed by Nation (Heatley, Nation, and Coxhead, 2002).

With regard to media and language learning, previous studies have confirmed that audiovisual programs play a positive role in the second language acquisition process. For instance, Inglese *et al.* (2007) studied how ESL learners perceive audiovisual interviews and found that visible author format eliminates the gap between the interview language and learners' linguistic ability. Similarly, William & Thorne (2000) shows the value of inter-lingual subtitling for SLA. Furthermore, language learners' motivations are also stimulated by media presentations. Bird (2005) showed that EAL learners have a more positive attitude toward language input via multimedia methods, and similar results were also reported in Bhrani (2011).

On the other hand, none of the previous studies have discussed media's potential benefits for foreign language lexical development. The audiovisual feature of TV dramas may provide extra semantic information for learners, which could be a beneficial factor. Meanwhile, learners' higher motivation of learning language through multimedia may also contribute to a better lexical acquisition. In addition, the lexical richness of TV dramas has rarely been tested in previous studies. All the above gaps in previous studies have given rise to the research questions of the current study.

3. Research Design
3.1 Research Questions

The following three questions are explored by the present study since previous research on TV sitcoms and vocabulary instruction are very rare.

(1) In terms of lexical range, can the TV sitcom *Friends* provide sufficient lexical input for Chinese EAL learners? If so, then which kind of English can better supported by these lexical inputs, English for General Purposes (EGP) or English for Academic Purposes (EAP)?

(2) With regard to the selected *Friends* corpus, what is its N-gram distribution pattern? Can the frequent lexical chunks in *Friends* provide sufficient lexical support for Chinese EAL learners, as suggested by Lewis (1993)?

(3) In General, can TV sitcom be regarded as an efficient tool for facilitating Chinese EAL learners' lexical acquisition? How can it be appropriately applied in TESL/TEFL programs in China?

3.2   Corpus Compilation

The corpus in the current study includes the scripts of ten episodes of *Friends*. The show *Friends* has a total of 236 episodes and 823,537 tokens. Such a large corpus can definitely cover a large range of English lexicons. Thus only a small portion of the corpus was selected in the current study. The focus of the present study is the efficiency of lexical input from the TV sitcom *Friends*. The total time for the ten episodes is around 200-220 minutes, approximately equal to five formal language classes in China (40 min/class). The ten episodes were selected randomly and generated 37,503 tokens in total. For detailed information of the ten episodes, see Appendix A.

3.3   Analysis Software

The programs used in the current study are Range and N-Gram Phrase Extractor. Range (Heatley et al, 2002) is able to compare a corpus with existing word lists such as the British National Corpus (BNC) and the Academic Writing List (AWL) (Coxhead, 2000). The analysis results of Range can be used to indicate the coverage of a text by certain word lists, to create word lists based on frequency and range, and to discover shared and unique vocabulary in several pieces of writing (Nation, 2004; Nation & Coxhead, 2001). N-gram is defined as a continuous sequence of n items in a given text. N-Gram Phrase Extractor is accessed from the Compleat Lexical Tutor website (http://www.lextutor.ca/). The tool can show the N-Gram patterns of the target corpus, which then suggests whether the target corpus covers frequent English grammatical chunks for learners' lexical development (Lewis, 1993, 2007).

3.4   Data Analysis Procedure
The scripts of the ten episodes of *Friends* were first converted to plain text format (txt) and then compiled a corpus named as the FRI corpus. Then the corpus was imported into Range and N-gram Phrase Extractor for analysis. There were two stages in the investigation: first, the lexical range of the FRI corpus was compared with both BNC and AWL to exam their overlaps. Then, the corpus' N-gram phrases were extracted via N-gram Phrase Extractor and the results were analyzed for their frequencies.

4.   Results
4.1   Lexical Richness of the FRI Corpus

Table 1 shows the comparison between the FRI corpus and the first 3000 word families in the Academic Word List. Word lists one to three are arranged according to the lexical frequencies of enlisted word families. For example, Word list one includes the most frequent 1000 word families in academic texts. From the results, it can be clearly observed that most tokens in the FRI corpus are in word list one, with a proportion of 75.5%. Similar results are found in the Word Types section in which tokens matching word list one represent 40.3%. By comparison, tokens in word lists two and three only took very small percentages of the FRI corpus: 5.1% of tokens fell into word list two and 0.8% into word list three. In terms of word families, 719 word families were found in word list one, followed by 405 in word list two and 108 in word list three. Another noticeable factor was that a considerable number of tokens were not in the three word lists. They comprised of 18.6% of the total tokens and their rate in the Word Types section was even higher, reaching 38.2%.

Table 1
Comparison between results of the FRI corpus and the Academic Word List

WORD LIST	TOKENS/%	TYPES/%	FAMILIES
One	28302/75.5	1327/40.3	719
Two	924/5.1	575/17.5	405
Three	296/0.8	133/ 4.0	108
Not in the lists	6981/18.6	1257/38.2	N/A
Total	37503	3292	1232

Table 2 provides information of the comparison between the FRI corpus and the 3000 most frequent word families in the British National Corpus. The results were similar to the results in Table 1. To be specific, 79.2% tokens were found in word list one, and they constituted 44.5 % of word types. All the 29,686 tokens in word list one formed 784 word families. The figures in word lists two and three resembled their counterparts in Table 1, with 3.9% tokens in word list two and 1.7% in word list three. The figures for the Word Types section were 17.3% and 9.0% respectively. It was worth noticing that the tokens in word list three in Table 3 had a higher rate than in Table 2, while the tokens that were not in the three word lists were quite numerous in Table 2 as well, with a total of 15.2% tokens, and 29.2% word types.

Table 2
Comparison between results of the FRI corpus and the British National Corpus

WORD LIST	TOKENS/%	TYPES/%	FAMILIES
One	29686/79.2	1466/44.5	784
Two	1451/ 3.9	568/17.3	414
Three	652/ 1.7	295/ 9.0	239
not in the lists	5714/15.2	963/29.2	N/A
Total	37503	3292	1437

To determine whether there is any statistical difference of data between Table 1 and Table 2, all the data were imported into SPSS 19.00 for further T-test analyses. Five pairs of data reached statistical significance ($p < 0.05$), namely tokens in word list one, tokens in word list two, word types in word list two,

word families in word list one and word families in word list two, as illustrated in Table 3.

Table 3
Independent T-test results between AWL and BNC

Type	AWL	BNC	*p* Value (Two-Tailed)
Tokens (Word List one)	75.5%	79.2%	.015
Word Types (Word List two)	17.5%	17.3%	.004
Word Families (Word List one)	719	784	.028
Word Families (Word List two)	405	414	.007

4.2   Results of N-Gram Analysis

Frequent N-Gram phrases were extracted from the FRI corpus and the top ten phases in different N-gram categories are shown in Tables 4, 5, and 6. Many colloquial lexical chunks can be found in the three-word and two-word strings, such as *oh my God*, *you know*, *I think*, and so on.

Table 4
Results of four-word strings of the FRI Corpus

4-wd strings: 35,559	
Repeated: 488 (1.37%)	
TTR: 488:1104 (1:2.26)	
Words: 1952 (5.48% of total)	
001.[9] The rest of the	006.[5] to talk to you
002.[8] End ___start life	007.[5] I want you to
003.[6] Chandler, Joey, and Ross's	008.[5] What do you mean
004.[6] The one with then	009.[5] to have a baby
005.[6] I don't want to	010.[5] of our lives set

Table 5
Results of three-word strings of the FRI Corpus

3-wd strings: 35,560	
Repeated: 1571 (4.42%)	
TTR: 1571:4240 (1:2.69)	
Words: 4713 (13.25% of total)	
001.[33] Oh my god	006.[14] I have to
002.[24] You know what	007.[13] No, No, No,
003.[19] I don't know	008.[13] What are you
004.[18] What do you	009.[12] The rest of
005.[16] Monica and Rachel's	010.[11] Have a baby

Table 6
Results of two-word strings of the FRI Corpus

2-wd strings: 35,561	
Repeated: 3637 (10.23%)	
TTR: 3637:15458 (1:4.25)	
Words: 7274 (20.45% of tot)	
001.[75] You know	006.[57] Do you

002.[66]	All right	007.[57]	This is
003.[65]	On the	008.[55]	In the
004.[64]	Are you	009.[53]	I think
005.[59]	To the	010.[52]	Have to

5.   Discussion of the Results

The above corpus analyses show the lexical range and frequent N-gram phrases in the FRI corpus. It should be noted that due to the lack of relevant studies in this perspective, the following discussions can only provide indirect implications.

The first research question tries to determine whether the sitcom *Friends* can provide sufficient lexical input for Chinese EAL learners. Results in Table 1 and Table 2 clearly showed that the ten episodes of *Friends* cover a large proportion of word families in both general and academic English. Considering that the ten episodes only last a total of 200-220 minutes, the wide lexical range of this TV sitcom is impressive and can be a valuable resource for Chinese EAL learners' lexical acquisition. Moreover, the lexical range of the FRI corpus displays depth as well since a considerable proportion of lexicons in the FRI corpus are not included in the first 3000 word families in both BNC and AWL. Thus the TV sitcom is a suitable learning resource for both intermediate level learners and advanced level learners. Furthermore, although the comparison between Table 2 and Table 3 indicated that there are differences in lexical range in terms of BNC and AWL, the FRI corpus generally suggested a similar pattern in both domains, and therefore the sitcom is a good learning research tool for both EGP and EAP. These results are in concordance with previous studies in Media and Language Learning such as Bahrani (2011), Bird (2005), Inglese et al. (2007), and Williams and Thorne (2000).

The second question concerns the appropriateness of applying the sitcom as language material for the "lexical approach" (Lewis 1993, 2007). Many N-gram phrases identified in the corpus were colloquial lexical chunks in daily communication, as shown in Tables 4, 5 and 6. This indicates that the sitcom is a good resource for students to imitate daily communication in English speaking countries. However, this claim should not be regarded as a strong claim since the N-gram lists in Tables 4, 5, and 6 also show that the N-gram distribution in the current FRI corpus was not systemic. The phrase inputs in the FRI corpus were merely based on word frequencies. Some very crucial argument structures such as the subjective use of verbs and the use of conjunctions did not appear in the N-gram lists. As discussed in Nation (2004), oral vocabulary acquisition is best realized in communicative task situations. The sitcom *Friends* should thus be used as an ancillary method but not as major teaching material.

Finally, as for research question three, the above analyses showed several beneficial factors of TV dramas for lexical acquisition. The FRI corpus can provide sufficient lexical input for EAL learners. Nevertheless, it is too risky to jump to the conclusion that TV dramas are beneficial for language learning. As discussed in previous studies such as Simard and Jean (2011), sufficient input does not necessarily lead to successful learner uptakes. In the current study, whether the TV sitcom is able to draw learners' attention to lexical input was unknown. It is possible that learners focus on the story of the TV sitcom and thus their lexical acquisition is affected as a result. In conclusion, the present results suggested that the TV sitcom *Friends* has a high possibility to be an

appropriate resource for EAL learners' lexicon instruction, but further studies are still needed.

6.   Conclusion

The findings of lexical range and N-Gram phrase in the FRI corpus have some interesting pedagogical implications for further EAL lexicon instruction research. The lexicons in the TV sitcom *Friends* cover most of the word families in both BNC and AWL, which suggests that instructors for both EGP and EAP may actively use this TV sitcom as valuable extra-curricular material to facilitate students' lexical acquisition process. As indicated in previous SLA theories such as Input Hypothesis (Krashen, 1983, 2003), comprehensive input is crucial for successful development in a second language. One problem for the inefficiency of EAL instruction in Asian countries is the lack of sufficient input outside the classroom (Cook, 2000), thus TV sitcoms can provide a possible solution to the above problem. Furthermore, with the development of internet technologies and online video websites, students will have easier access to a vast collection of TV sitcoms in the near future. To conclude, using TV sitcoms as a foreign language instruction tool has the possibility to become a prominent research topic in the near future and more studies in this area are required.

Meanwhile, several limitations of the present study should be considered. The corpus only includes one popular US sitcom in China, which is somewhat limited. The lexical range and N-Gram phase in other TV drama genres (e.g. police procedurals and medical dramas) are not investigated in the current study. Thus it is not plausible to claim that all English TV dramas are effective for English lexical acquisition. Similarly, the effect of TV dramas on other aspects of language learning, such as syntactic rules and pragmatics are not considered, which leaves room for further studies.

References

Bahrani, T. 2011. The role of audiovisual mass media news in language learning. *English Language Teaching 4* (2): 260-266.

Bennett, G. 2010. *Using Corpora in the language learning classroom: Corpus linguistics for teachers*. Ann Arbor: Michigan University Press.

Bird, S. A. 2005. Language learning edutainment: mixing motives in digital resources. *RECL 36* (3): 311-339.

Blake, J. 2008. *Brave new digital classroom*. Washington D.C.: Georgetown University Press.

Block, J. 1995. A contrarian view of the five-factor model of personality description. *Psychological Bulletin 117*: 187–215.

Celce-Murcia, M. 2001. Language teaching approaches: an overview. In *Teaching English as a second or foreign language,* ed. Marianne Celce-Murcia, 3-11. Boston: Heinle Cengage.

Chen, J. 2010. *A corpus-based study of affect in Friends*. Unpublished MA Thesis, Shandong University, China.

Cook, V. 2000. *Second language learning and language teaching*. Beijing: Foreign Language Teaching and Research Press.

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

Gavioli, L., and Aston, G. 2001. Enriching reality: language corpora in language pedagogy. *ELT Journal 55* (3): 238-246.

Gass, S. M. 1988. Second language vocabulary acquisition. *Annual Review of Applied Linguistics 9*: 92-106.

Hall G. 1994. Review of the lexical approach: *The state of ELT and a way forward*, by Michael Lewis. *ELT Journal 44*: 48.

Heatley, A., Nation, I.S.P. and Coxhead, A. 2002. RANGE and FREQUENCY programs. Retrieved from: http://www.vuw.ac.nz/lals/staff/Paul_Nation

Hu, L. J. 2007. *A corpus-based study of hedges in Friends*. Unpublished MA Thesis, Central University of Science and Technology, China.

Huang, L.-S. 2008. Using guided, corpus-aided discovery learning to generate active learning. *English Teaching Forum 46* (4): 20-27.

Inglese, T., Mayer, R. E. & Rigotti, F. 2007. Using audiovisual TV interviews to create visible authors that reduce the learning gap between native and non-native language speakers. *Learning and Instruction 17*: 67-77.

John, T. 1997. Contexts: the background, development and trialing of a concordance-based CALL program. In *Teaching and language corpora,* eds. Anne Wichmann, Steven Fligelstone, Tony McEnery and Gerry Knowles, 100-115. London: Longman.

Krashen, S. D. & Terrell, T.D. 1983. *The natural approach: Language acquisition in the classroom*. San Francisco: The Alemany Press.

Krashen, S. D. 1985. *The input hypothesis: Issues and implications*. New York: Longman.

Krashen, S. D. 2003. *Explorations in language acquisition and use*. Portsmouth: NH: Heinemann.

Laufer, B., and Nation, I.S.P. 1995. Lexical richness in L2 written production: can it be measured? *Applied Linguistics 16* (3): 307-322.

Lewis, M. 1993. *The Lexical approach: The state of ELT and the way forward*. London: Language Teaching Publications.

Lewis, M. 1997. *Implementing the lexical approach: Putting theory into practice*. London: Language Teaching Publications.

Lewis, M. (ed.) 2000. *Teaching collocation: Further developments in the lexical approach*. Hovem England: Language Teaching Publications.

Nation, I.S.P. 1993. Vocabulary size, growth and use. In *The bilingual lexicon*, eds. Robert Schreuder & Bert Weltens, 115-134. Amsterdam/Philadelphia: John Benjamins.

Nation, I.S.P. 2004. A study of the most frequent word families in the British National Corpus. In *Vocabulary in a second language: Selection, acquisition and testing*, eds. Paul Bogaards, Batia Laufer and Batia Laufer-Dvorkin, 3-13. Amsterdam/Philadelphia: John Benjamins.

Nation, I.S.P. & Coxhead, A. 2001. The specialized vocabulary of English for academic purposes. In *Research perspectives on English for academic purposes*, eds. John Flowerdew and Matthew Peacock, 252-267. Cambridge: Cambridge University Press.

Reppen, R. 2010. *Using Corpora in the Language Classroom*. Cambridge: Cambridge University Press.

Simard, D., and Jean, G. 2011. An exploration of L2 teacher's use of pedagogical interventions devised to draw L2 learners' attention to form. *Language Learning 61* (3): 759-785.

Westen, D. 1996. A model and a method for uncovering the nomothetic from the idiographic: an alternative to the five-factor model? *Journal of Research in Personality 30*: 400–413.

Williams, H. & Thorne, D. 2000. The value of teletext subtitling as a medium for language learning. *System 28* (2): 217-228.

Ye J. 2005. *A corpus-based study of idioms in Friends*. Unpublished MA Thesis, Central University of Science and Technology, China.

Appendix A

Episodes of *Friends* Used for the FRI Corpus

Friends - 1x05 - The East German Laundry Detergent
Friends - 2x03 - Heckles Dies
Friends - 3x12 - All the Jealousy
Friends - 4x09 - They're Gonna PARTY
Friends - 5x10 - To the Inappropriate Sister

Friends - 6x06 - To On the Last Night
Friends - 7x03 - To Phoebe's Cookies
Friends - 8x03 – The One Where Rachel Tells
Friends - 9x14 - To the Blind Dates
Friends - 10x02 - Ross Is Fine