

---

# TV DRAMA AND LEXIS: A CORPUS-BASED LEXICAL DIVERSITY STUDY

Sibo Chen

(*Simon Fraser University, Canada*)

## **Abstract**

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 use English TV drama as an effective method for EAL lexical instruction. This article reports a corpus-based study relating to the popular US TV sitcom *Friends*. The research question is 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 indicates that English TV sitcoms may serve as an effective tool for EAL lexical instructions.

## **1. Introduction**

With the enhancement of technology, especially the rapid 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 21<sup>st</sup> century is developing towards a digital version. Stimulated by Internet technologies, 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 search 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 benefits of multimedia programs for SLA, such as Bahrani (2011), Bird (2005), Inglese, Mayer & Rigotti (2007), Webb & Rodgers (2009) and William & Thorne (2000), English TV dramas have rarely been studied from a lexical instruction perspective. Bearing in mind this gap in previous literature, the present study focuses on the lexical diversity of English TV dramas. The study selected ten episodes from

the US TV sitcom *Friends* as the research data, since *Friends* is generally regarded as one of the most popular English TV dramas in China due to its early introduction to Chinese audience in the late 1990s. The sitcom's popularity and lively scripts have attracted many Chinese applied linguists. Previous studies concerning *Friends* in language education have been carried out from many perspectives including collocation and idioms (Ye 2005), hedges (Hu 2007), and affective metaphor (Chen 2011). The current study, by comparison, incorporates the corpus linguistics method and explores the lexical diversity of the sitcom, focusing on whether the sitcom can provide sufficient recourse for EAL learners' lexical development.

## 2. Literature review

### 2.1. *Lexical instruction and the lexical approach*

Learners' knowledge of the lexicon of a foreign language has a fundamental influence on their performance in that language. As suggested by previous studies, 95% to 98% lexical coverage is needed for an adequate comprehension of English texts and TV programs (Laufer 1989; Hu & Nation 2000; Nation 2006). Unlike native English speakers who learn a large quantity of lexical items over many years, EAL learners often face the challenge of grasping a large stock of new terms in a relatively short time, which makes vocabulary instruction particularly challenging for many TESL/TEFL programs (Folse 2011).

Despite the significance of lexicon acquisition, it cannot be effectively supported by current mainstream EAL pedagogies. Traditional teaching methods (e.g. the Grammar-Translation Approach and the Audio-Lingual Method) apply the drill-practice method for lexicon memorization explicitly, which is reported as detrimental for learners' initiative as well as enthusiasm in language learning (Celce-Murcia 2001; Cook 2000). In recent years, the widespread application of Communicative Language Teaching (CLT) has relegated vocabulary instruction to a secondary position, and learners are expected to expand their lexical inventory by implicit and incidental learning during communicative interactions with the target language (Decarrico 2001). Nevertheless, several studies, such as Gass (1988) and Nation (1993), show the ineffectiveness of CLT on lexical acquisition. According to these studies, a learner's lexicon inventory is somewhat limited despite being involved in CLT programs for an extended period. In general an inventory of 3000 words is a crucial threshold in a learner's second language acquisition process (Nation 1993). After this threshold, L2 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, many methods have been developed for teaching L2 vocabulary, and one of the most influential methods among them is Lewis's Lexical Approach (LA) (Lewis 1993, 1997). The basic concept of LA is that vocabulary instruction should be prioritized with respect to grammar instruction. 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. In summary, lessons in LA are supposed to focus on fixed expressions that occur frequently in the target language's daily conversations (Lewis 1993).

LA has received contradictory reviews since its appearance. Hall (1994), in a review of Lewis's work, emphasizes the approach's ambitious perspective and its innovational ideas, whereas the approach has also been criticized for its lack of theoretical grounds and inadequacy of teaching personality structures (e.g. Block 1995; Westen 1996). As a response to the critics proposed by other scholars, Lewis published follow-up works on the application of LA (Lewis 1997: 2000), in which he exemplifies how the conceptions of LA can be effectively applied in real language classrooms. However, one critical challenge for the implementation of LA, as admitted in Lewis's works, is the selection of appropriate course material. Many traditional EAL reading materials are appropriate for LA since they lack the essential lexical diversity feature for LA instruction.

## 2.2. *Corpus-based lexical research and previous studies in media and language instruction*

Corpus linguistics is becoming a major field in SLA research along with the advancement of corpus compilation and corpus analysis software (Bennett 2010; Reppen 2010; Huang 2008). By using large quantities 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, with so many electronic corpora available online, he/she can examine a particular word, compare different texts, and 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 (DDL) (Gavioli & Aston 2001), in which students act as "language detectives" to actively participate in the discovery of language patterns.

With regard 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). One key concept in numerous studies by Nation is lexical diversity, which describes the range of words in a target text. In the current study, the lexical diversity of the scripts of the TV sitcom *Friends* is investigated with the application of the corpus software tool developed by Nation and his colleagues (Heatley, Nation & Coxhead 2002).

In terms of media and language learning, previous studies have confirmed that multimedia 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) show the value of interlingual 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).

Nevertheless, few of the previous studies have discussed the media's potential benefits for foreign language lexical development. Although Webb & Rodgers (2009) have provided a comprehensive examination of lexicon coverage in English movies, the lexicon coverage in English TV dramas has rarely been studied. In addition, the audiovisual feature of multimedia may provide comprehensive input for learners, and learners' higher motivation of learning language through multimedia may also contribute to a better lexical acquisition. To sum up, all the gaps in previous studies have contributed to the formulation of research questions in the present study.

### 3. Research design

#### 3.1. Research Questions

The following three questions are explored in the present study:

(1) In terms of lexical diversity, can the TV sitcom *Friends* provide sufficient lexical input for Chinese EAL learners? If the answer is yes, then which type of English program can be better supported by the sitcom, English for General Purposes (EGP) or English for Academic Purposes (EAP)?

(2) What is the N-gram distribution pattern of the selected *Friends* corpus? Can frequent lexical chunks in *Friends* provide sufficient lexical support for Chinese EAL learners, as defined by Lewis (1993)?

(3) In general, can TV sitcoms be regarded as an efficient tool for facilitating Chinese EAL learners' lexical acquisition, and how can it be appropriately applied to TESL/TEFL programs in China?

#### 3.2. Corpus compilation

The sitcom *Friends* is made up of a total of 236 episodes and 823,537 tokens which, because of its size, can definitely cover a wide range of English lexicons. Thus the focus of the present study is the efficacy of lexical input from *Friends*. Ten episodes of *Friends* were randomly selected, and the total time of these episodes was approximately 200-220 minutes, equal to five formal language lectures in China (40 minutes per class). The scripts of the selected episodes generated 37,503 tokens in total. These scripts were first converted to plain text format (txt) and then compiled into a corpus named the FRI corpus. Table 1 provides detailed information of the selected episodes.

---

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

---

**Table 1.** Episodes of *Friends* used for the FRI corpus

### 3.3. Data analysis procedure

The corpus analysis programs used in the present study were Range and N-Gram Phrase Extractor. Range is developed by researchers at Victoria University of Wellington (Heatley *et al.* 2002) and it 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, as well as 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. The 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, 1997).

There were two stages in the present study. First, the FRI corpus was imported into Range and its tokens were compared with both BNC and AWL to investigate lexical overlaps between the two word lists and the FRI corpus. The independent sample t-test was used to determine whether there was a significant difference between the coverage of BNC and AWL in the FRI corpus. Second, using the N-gram Phrase Extractor, the FRI corpus' N-gram phrases were extracted and the N-gram frequency lists were further analysed.

## 4. Results

### 4.1. Lexical diversity of the FRI corpus

Table 2 shows the comparison between the FRI corpus and the first 3,000 word families in the Academic Word List. Word lists one to three are arranged according to the lexical frequencies of listed word families. For instance, Word List One includes the most frequent 1,000 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 represent 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 18.6% of the total tokens and their rate in the Word Types section was even higher, reaching 38.2%.

WORD LIST	TOKENS/%	TYPES/%	FAMILIES
One	28,302 = 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	37,503	3292	1232

**Table 2.** Comparison between results of the FRI corpus and the Academic Word List

Table 3 provides information of the comparison between the FRI corpus and the 3,000 most frequent word families in the British National Corpus. The results were similar to the results in Table 2. 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 2, 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 is worth noticing that the tokens in Word List Three in Table 4 had a higher rate than in Table 3, while the tokens that were not in the three word lists were quite numerous in Table 3 as well, with a total of 15.2% tokens, and 29.2% word types.

WORD LIST	TOKENS/%	TYPES/%	FAMILIES
One	29,686 = 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	37,503	3292	1437

**Table 3.** Comparison between results of the FRI corpus and the British National Corpus

To determine whether there is any statistical difference of data between Table 2 and Table 3, 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 4.

Type	AWL	BNC	<i>p</i> 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

**Table 4.** Independent T-test results between AWL and BNC

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*, or *I think*.

---

4-word 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 them	009.[5] to have a baby
005.[6] I don’t want to	010.[5] of our lives set

---

**Table 5.** Results of four-word strings of the FRI Corpus

---

3-word 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 three-word strings of the FRI Corpus

---

2-word strings: 35,561	
Repeated: 3637 (10.23%)	
TTR: 3637:15458 (1:4.25)	
Words: 7274 (20.45% of total)	

---

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

---

**Table 7.** Results of two-word strings of the FRI Corpus

## 5. Discussion of the results

The analysis presented in section 4 above shows 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 tentative implications.

The first research question tries to determine whether the sitcom *Friends* can provide sufficient lexical input for Chinese EAL learners. Results in Table 2 and Table 3 clearly show that the ten episodes 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 lexical items in the FRI corpus are not included in the first 3,000 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 3 and Table 4 indicates that there are differences in lexical range in terms of BNC and AWL, the FRI corpus generally suggests 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 (Bahrani 2011; Bird 2005; Inglese *et al.* 2007; Webb & Rodgers 2009; Williams & Thorne 2000).

The second question concerns the appropriateness of applying the sitcom as language material for LA. Many N-gram phrases identified in the corpus were colloquial lexical chunks in daily communication, as shown in Tables 4, 5 and 6. These results indicate 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 show 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 & 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 is 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 of being 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 lexicon in *Friends* covers 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 the students' lexical acquisition process. As indicated in previous SLA theories such as Input Hypothesis (Krashen 1985, 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.

## 7. Acknowledgement

I would like to thank Dr. Li-Shih Huang at the Department of Linguistics, University of Victoria for her comments on earlier drafts of this paper. All remaining errors are my own.

## 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 DC: 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 M. Celce-Murcia (ed.), *Teaching English as a Second or Foreign Language*. Boston: Heinle Cengage: 3-11.
- 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.
- DeCarrico J. S. 2001. Vocabulary learning and teaching. In M. Celce-Murcia (ed.), *Teaching English as a Second or Foreign Language*. Boston: Heinle Cengage: 285-300.
- Folse K. 2011. Applying L2 lexical research findings in ESL teaching. *TESOL Quarterly* 45/2: 362-369.
- Gavioli L. & G. Aston 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., I.S.P. Nation & A. Coxhead 2002. *Range and Frequency* programs. Available online at: [http://www.vuw.ac.nz/lals/staff/Paul\\_Nation](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.
- Hu M. & I.S.P. Nation 2000. Vocabulary density and reading comprehension. *Reading in a Foreign Language* 13: 403-430.
- Huang L.-S. 2008. Using guided, corpus-aided discovery learning to generate active learning. *English Teaching Forum* 46/4: 20-27.
- Inglese T., R.E. Mayer & F. Rigotti 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 A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (eds.), *Teaching and Language Corpora*. London: Longman: 100-115.
- 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. 1989. What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (eds.), *Special Language: From Humans Thinking to Thinking Machines*. Bristol: Multilingual Matters: 316-323.
- Laufer B. & I.S.P. Nation 1995. Lexical diversity 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. 2000. Language in the lexical approach. In M. Lewis (ed.) *Teaching Collocation: Further Developments in the Lexical Approach*. Hove, England: Language Teaching Publications.
- Nation I. S. P. 1993. Vocabulary size, growth and use. In R. Schreuder & B. Weltens (eds.), *The Bilingual Lexicon*. Amsterdam/Philadelphia: John Benjamins: 115-134.
- Nation I. S. P. 2004. A study of the most frequent word families in the British National Corpus. In P. Bogaards, B. Laufer & B. Laufer-Dvorkin (eds.), *Vocabulary in a Second Language: Selection, Acquisition and Testing*. Amsterdam/Philadelphia: John Benjamins: 3-13.
- Nation I. S. P. 2006. How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review* 63: 59-82.
- Nation I.S.P. & A. Coxhead 2001. The specialized vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (eds.), *Research Perspectives on English for Academic Purposes*. Cambridge: Cambridge University Press: 252-267.
- Reppen R. 2010. *Using Corpora in the Language Classroom*. Cambridge: Cambridge University Press.
- Simard D. & G. Jean 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.
- Webb S. & M.P.H. Rodgers 2009. The lexical coverage of movies. *Applied Linguistics* 30/3: 407-427.
- 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. & D. Thorne 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.