Lexical Bundles in the Korean EFL Teacher Talk Corpus: A Comparison Between Non-native and Native English Teachers

Ye-Eun Kwon
Kunsan National University
Eun-Joo Lee
Ewha Womans University

The present study investigates the use of lexical bundles in Korean EFL teacher talk. In particular, it examines the quantitative and qualitative features of lexical bundles in a native and non-native teachers' classroom talk corpus. The corpus of the study consists of a total of 247,398 words complied through 62 hours of recording of EFL classes. The key lexical bundles of the study were extracted using *WordSmith 6.0* to conduct frequency, functional and structural analyses. The results of the frequency analysis showed the non-native English teachers' repeated use of a limited number of lexical bundles which inflated the total number of usages. The functional analysis of the lexical bundles showed that the non-native English teachers relied heavily on a narrow range of functional categories, e.g., stance expressions and referential expressions. Further, the results of the functional analysis clearly indicate that the EFL teacher corpus possesses certain repetitive tendencies in both conversational and academic discourse. The examination of the over-, exclusive, and under-use of lexical bundles demonstrates that not only quantitative but also qualitative differences exist in lexical bundle preferences between the native and the non-native teachers.

Keywords: lexical bundles, corpus analysis, teacher talk

INTRODUCTION

Control of lexical bundles is a key component of both oral and written discourse as the effective use of the multi-word expressions contributes to successful and fluent linguistic production (Biber & Barbieri, 2007; Hyland, 2008). Thus, it is only natural that second language (L2) studies on lexical bundles reported that lexical bundles are useful as a production strategy for both beginning and advanced learners (Boers, Eyckmans, Kappel, Stengers & Demecheleer, 2006; Conklin & Schmitt, 2008; Underwood, Schmitt & Galpin, 2004; Wood, 2006; Wray, 2002; Wray & Fitzpatrick, 2008). These studies have shown that lexical bundles are conventionalized in a certain register, and thus understanding characteristics of them will be beneficial for L2 speakers as a means to improve their speaking skills. Studies such as Biber and Barbieri (2007), Biber, Conrad and Cortes (2004), Cortes (2004), Granger (1998), Howarth (1998), and Tribble (2011) revealed that certain types of lexical bundles are more prevalent in a specific pragmatic register. Cortes (2004), for example, examined what types of lexical bundles were more prevalent in a specific register. Cortes compared frequently occurring lexical bundles used by experts and non-experts in the field of history and biology and found out that the experts certainly used a different set of lexical bundles. Cortes concluded that learning a certain set of lexical bundles would help non-experts group to improve their writing skill in a specific field. Previous studies on lexical bundles in the classroom settings investigated different linguistic features across various disciplines, modes, and languages (Baker, 2009; Kemppanen, 2004; Ku & Yang, 1999; Mauranen, 2006; McEnery, 2009; Seales, Charteris-Black & Ziebland, 2006). Doyle and Hong (2009), for example, examined disciplinary differences in teacher talk using the SCoRe (Singapore Corpus of Research in Education) collected from Singapore primary and secondary school classes taught in English. In their pioneering research on lexical bundles, Biber and Conrad (1999) and Biber et al. (2004) identified the differences between spoken and written discourse, academic spoken discourse and conversation, and novice and professional

writing.

Using lexical bundles as the unit of language data analysis has been proven to be an effective strategy for finding features in large datasets as it combines the quantitative element of frequency information with a qualitative element (Biber et al., 2004; Biber, Johansson, Leech, Conrad, & Finegan, 1999). Even though many applied linguists have emphasized the role of lexical bundles in the production and comprehension of language, relatively little attention has been paid to this line of research (Granger & Paquot, 2010; Wray, 2002). In addition, as De Cock (2004) noted, only a few studies have examined the characteristics of the learner spoken corpus. No attempt has been made to investigate EFL teacher talk by compiling an EFL teacher corpus and analyzing teacher talk using lexical bundles as a unit of analysis. Therefore, examining lexical bundles in non-native and native English teacher corpora is a new attempt to identify the phraseological patterns used by advanced English speakers and to compare these uses with their counterparts in a corpus of native English teachers. Investigating the authentic EFL teacher talk by noting the frequency of lexical bundles and comparing lexical bundle lists for non-native and native English teachers would be a first step toward practical suggestions for non-native teachers to improve their oral proficiency and for educators to improve teacher training programs and materials.

The present study analyzed the use of lexical bundles in the Korean EFL teacher talk corpus. In particular, the study investigates both quantitative and qualitative features on lexical bundle usages by comparing the non-native EFL teacher (NNET) corpus with the native EFL teacher (NET) corpus. The study addresses the following research questions:

1. What are the functional and structural features of EFL teachers' lexical bundles? What are the similarities and/or differences in the functional and structural features of lexical bundles between the NNET corpus and the NET corpus?
2. Which types of lexical bundles are over-, exclusive, and under-used in the NNET corpus? What are the qualitative characteristics of over-, exclusive and under-use of lexical bundles?

METHOD

Corpus of the Study

Participants

The corpus for the study was collected from four Korean and five native university EFL instructors. All the NNETs were teaching general English classes at three different universities in Seoul at the time of the data collection. For comparative purposes, five native teachers who were teaching at two different universities in Seoul participated in the study. Table 1 summarizes demographic information of the EFL teachers who participated in the study.

TABLE 1
Demographic Information of the EFL Teachers

Teacher	Nationality	Gender	Age	Education Level	Major	TEE[1] Experiences	TOEIC Scores
NNET-1	Korean	Female	38	Ph.D.	English education	4 years	955
NNET-2	Korean	Female	34	M.A.	Education	6 years	990
NNET-3	Korean	Female	34	M.A.	Education	3 years	990
NNET-4	Korean	Female	42	M.A.	English education	4 years	975
NET-1	British	Male	35	M.A.	Hospitality management	8 years	none
NET-2	Canadian	Male	34	M.A.	International Policy	7.5 years	none
NET-3	Canadian	Female	45	M.A.	Business	6 years	none
NET-4	American	Female	28	M.A.	International Public affairs	4 years	none
NET-5	American	Female	35	M.A.	Linguistics	9 years	none

As displayed in Table 1, NNETs are Korean and NETs are from North America including the U.S. and Canada, except for one British participant.

---

[1] TEE = Teaching English through English

The EFL teachers' age ranged from the late twenties to mid-forties. The Korean EFL teachers held graduate level degrees in English education or general education while the native EFL teachers, graduate level degrees in fields other than English or education. The participants' teaching experience ranged from three to nine years and the Korean EFL teachers had on average slightly less experience teaching English than native EFL teachers. The Korean EFL teachers' standardized English scores ranged from 955 to 990 in TOEIC.

NNET and NET Corpus

Lectures from participants were audio-recorded and transcribed to build the teacher talk corpus for the study. As the study focused on the characteristics of teacher talk, students' talk was not included in the corpus although some of their talk was transcribed. The compiled teacher talk corpus was uploaded onto www.lextutor.ca for storage, distribution, and future uses. The EFL teacher talk corpus consists of two sub-corpora: non-native EFL teacher (NNET) and native English teacher (NET) corpus. The NNET corpus includes lectures from four Korean teachers of English who studied abroad after they had passed the critical period of language learning. This late exposure is one of the selection criteria for NNETs as it reflects the reality where the majority of Korean teachers of English have not been exposed to English-speaking environments before adolescence.

Table 2 shows the composition of the EFL teacher talk corpus. As the table reveals, the corpus contains 247,398 words (123,122 words for NNET corpus and 124,275 words for NET corpus) representing approximately 62 hours of classroom teaching.
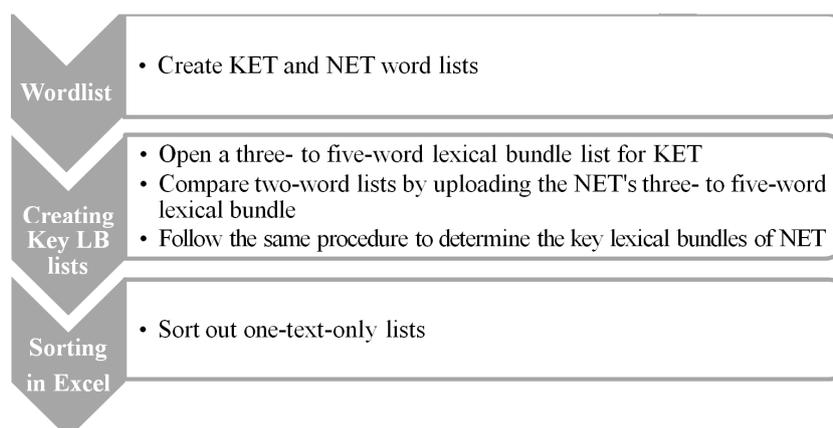
TABLE 2
Composition of EFL Teacher Corpus

Corpus	Word count	Time
NNET	123,122	1,940 min. (about 32 hrs)
NET	124,276	1,786 min. (about 30 hrs)
Total	247,398	3,726 min. (about 62 hrs)

ANALYSIS

Identification of Lexical Bundles

EFL teachers' lexical bundles of the study were extracted using *WordSmith 6.0* (Scott, 2012) which identifies lexical patterns up to a span of eight words. The concordance program not only shows the frequency of each lexical bundle but also calculates the percentage of lexical bundle occupies within the corpus and reports the number of texts in which the lexical bundle appears. Lexical bundles found in only one text file, which meant only one participant, were excluded in each corpus (NNET vs. NET) to prevent idiosyncratic speaker uses and the influence of different topics.

The study attempts to locate any typical and atypical patterns of lexical bundle usage of NNET in comparison with NET. Figure 1 displays summarized steps identifying lexical bundles for the study. In order to look into the different lexical bundle usages of NNET and NET, the study focused on the key lexical bundles, which are extracted by calculating a *keyness* score by using a log-likelihood value to rank lexical bundles in a target corpus and a comparing corpus (Scott, 2012). Key lexical bundles in a target corpus, which have higher log-likelihood values, are differentiated phrases from the comparing corpus.

Wordlist	• Create KET and NET word lists
Creating Key LB lists	• Open a three- to five-word lexical bundle list for KET • Compare two-word lists by uploading the NET's three- to five-word lexical bundle • Follow the same procedure to determine the key lexical bundles of NET
Sorting in Excel	• Sort out one-text-only lists

FIGURE 1
Key Lexical Bundle Identification Steps

To extract a list of key words or key lexical bundles, p-value was set to limit the number of key words. The p-value was set at .0001 to achieve a limited number of lexical bundles for practical reasons (Simpson & Ellis, 2010). Given the size of the corpus and the lexical bundle uses of NETs, the optimal frequency cut off point was set at five. After key lexical bundle lists had been extracted, a sorting step to lessen the influence of idiosyncratic usage and different topics was needed. After these steps, functional and structural analyses were conducted based on the taxonomy used by Simpson and Ellis (2010) and Cortes (2008).

Functional and Structural Taxonomy of Lexical Bundles

The functional and structural analysis of the lexical bundles was conducted based on the extracted key lexical bundles. The functional analysis of the study adopted the taxonomy by Simpson and Ellis (2010). Table 3 shows the taxonomy used in the study.

TABLE 3

Functional Taxonomy (Adapted from Simpson & Ellis, 2010, pp. 498-502)

Functional Categories	Subcategories	Examples
Referential expressions	Specification of features Identification and focus	based on (a/the), (as) part of (a/the), (an/the) example of (a) how many of you, one of these, and this is
	Contrast and comparison	and the same, between the two, is much more
	Time, place and situations	*a* and *b*, (at) the end (of) (the), at this point
	Vagueness markers	and so on, and so on so, blah blah blah
Stance expressions	Hedges	(more) likely to (be), (it/there) may be, a kind of, it could be
	Epistemic stance	according to the, I think this is, to figure out (what), what do you mean
	Obligations and directive	do you want (me) (to), I want you to, you don't need to,
	Expressions of ability and possibility	can be used (to), (gonna) be able (to), that you can, so you can (see)
	Evaluation	the importance of
	Intention/desire, prediction	I just wanted to, if you wanna, if you were (to)
Discourse organizing expressions	Topic introduction and focus Topic elaboration	come back to, I'm talking about, we talk(ed) about, what I'm saying, you're talking about for example, any questions about, come up with (a), I mean if (you), because it is
Others	Politeness Simple request Simple report Feedback	thank you very (much), good afternoon what did you do I said to her no no no (no), oh my god

As illustrated in Table 3, the functional category consists of four main sub-categories. The first category, *referential expressions*, means lexical bundles that refer to specific items, concepts, or expressions with a specified quantity. The second category is *stance expressions*, which shows attitudes or intentions of the speakers. The third category, *discourse organizing*

expressions, refers to discourse markers that connect sentences or paragraphs that introduce or elaborate on topics. The last category, *others*, refers to other expressions, such as greetings to show politeness, and other simple questioning and reporting expressions. The original contexts were consulted to ensure that the primary functions of each lexical bundle are maintained.

For the structural analysis of the key lexical bundles, the study employed the taxonomy of Cortes (2008) as shown in Table 4.

TABLE 4
Structural Taxonomy (Adapted from Cortes, 2008, p. 59)

Structural category	Examples
NP	one more time, something like that, the correct answer, your
(connector +) NP	group members
NP + PP	an example of, the meaning of
VP	are talking about, want you to do, can say that, check the answers, don't need to, fill in the blanks, going to give you, have any questions
PP	on the board, on the exam, to the next page
(connector +) Cl	are you finished, if you look at, so how would you, what I want you to, you have to use, and let's move to, are you ready to, do you have to, how can you say, if you look at the, I'm going to give you, let's move to, let's talk about, so it means, so the answer is, what do you mean by
Others	how about this one, and so on

As Table 4 displays, *NP* refers to lexical bundles categorized into noun phrases. Likewise, *VP* refers to verb phrases, *PP*, prepositional phrases, and *Cl,* clauses. *Others* encompassed lexical bundles which consisted of phrases used as independent sentences including adverbial phrases. The structural categories of taxonomy were determined based on the first word of lexical bundles. For example, a lexical bundle of *on the exam* was categorized into *PP*.

RESULTS AND DISCUSSION

Frequency Analysis of Lexical Bundles

The results of the frequency analysis showed that the frequency of the types of key lexical bundles was 364 in NNET and 220 in NET. After sorting the key lexical bundles, the numbers of NNET and NET decreased to 311 in NNET and 116 in NET.

Table 5 shows the top-10 key lexical bundles after excluding topic-or textbook-influenced lexical bundles. It includes additional information such as the frequency information and the percentage of key lexical bundles in each corpus.

To minimize the effects of topics or an individual speaker, topic-specific or only one speaker's use of key lexical bundles was excluded from the study. Table 5 shows the top-10 key lexical bundles after excluding topic-or textbook-influenced lexical bundles. It includes additional information such as the frequency information and the percentage of key lexical bundles in each corpus.

TABLE 5
Top-10 Key Lexical Bundles after Sorting[2]

NNET				NET		
Percent (Percent in NET)	Token Frequency (Frequency in NET)	Lexical Bundles	Rank	Lexical Bundles	Token Frequency (Frequency in NNET)	Percent (Percent in NNET)
35(8)	383(95)	you have to	**1**	do you think	155 (68)	14(6)
10(6)	113(71)	okay, let's	**2**	you have to	95(383)	8(35)
7	80(6)	let's move	**3**	want you to	74(18)	7(2)
7	77(0)	you can see	**4**	what did you	73(31)	6(3)
7(2)	73(27)	okay how about	**5**	look at the	73(173)	6(16)
6	62(7)	you look at the	**6**	I want you	68(15)	6(1)
5	59(0)	listen to the	**7**	I want you to	63(15)	6(1)
5	57(4)	move on to	**8**	you can say	63(132)	6(12)
5	56(5)	if you look at the	**9**	can you read	39(6)	3(0)
5	53(0)	can you please	**10**	a look at	32(6)	3(0)

As shown in Table 5, the three most frequently employed lexical bundles in the NNET corpus were *you have to, okay let's,* and *let's move* comprising 52%; in the NET corpus, *do you think*, *you have to* and *want you to* comprising 29%. NNETs employed a certain type of lexical bundles significantly more or less than NETs did regardless of the topic or an individual speaker. In other words, the frequency disparity of the top-10 key lexical bundles between the NNET corpus and the NET corpus showed that a few overly repeated lexical bundles in NNET dominated the percentage of NNET's lexical bundle list.

---

[2] The column of 'Frequency in NET' included cells marked as 0, which actually refer to lower frequency bundles under 4. The minimum frequency was set as 5 and WordSmith did not mark frequencies under 4 on the key lexical bundle results. The last column of 'Percent in NET' has blank cells, of which percentages of lexical bundles of less than 1% were not marked in the WordSmith.

Functional and Structural Analysis

The functional features of the lexical bundles were analyzed based on the functional taxonomy. Table 6 provides the results of the functional analysis of the lexical bundles.

TABLE 6
Frequency of Key Lexical Bundles by Functional Category

Functional category	NNET		NET	
	Freq.	Percent	Freq.	Percent
A. Referential expressions	87	28.0	31	24.6
(1) Specification of features	13	4.2	5	7.14
(2) Identification and focus	47	15.1	18	14.3
(3) Contrast and comparison	2	0.6	0	0
(4) Time, place and situations	25	8.0	7	5.6
(5) Vagueness markers	0	0	1	0.8
B. Stance expressions	170	54.7	73	57.9
(1) Hedges	2	0.6	2	1.59
(2) Epistemic stance	22	7.1	13	10.3
(3) Obligation and directive	74	23.8	29	23
(4) Expressions of ability and possibility	24	7.72	8	6.35
(5) Evaluation	7	2.25	2	1.59
(6) Intention/desire, prediction	41	13.2	19	15.1
C. Discourse organizing expressions	42	13.5	16	12.7
(1) Topic introduction and focus	37	11.9	7	5.56
(2) Topic elaboration	5	1.61	9	7.14
D. Others (Small talks)	12	3.54	6	4.76
(1) Politeness	0	0	0	0
(2) Simple request	6	1.93	5	3.97
(3) Simple report	0	0	0	0
(4) Feedback	6	1.93	1	0.79
Total	311	100	126	100

As shown in Table 6, the most frequently employed key lexical bundles were *stance expressions* in both NNET and NET corpora. Among the top 10 key lexical bundles in Table 5, lexical bundles categorized into *stance expressions* were most frequently used, comprising 6 out of 10 for NNET and 9 out of 10 for NET. The result shows that the lexical bundles with the most significant disparity in frequency were similarly distributed over all sub-categories in both NNET and NET corpus. NNETs and NETs employed a similar number of lexical bundles functionally among key lexical bundles.

As Biber and Barbieri (2007) noted, lexical bundles as *stance expressions* occur more frequently in conversations. The high frequency of the stance expressions shows that the teacher talk corpus has the features of the conversational discourse. The second most frequently used key lexical bundles were *referential expressions*, which generally occurred more frequently in academic discourse (Biber et al., 1999). This may reflect the fact that teacher talk used in general English courses also has the characteristics of academic discourse.

The structural analysis of key lexical bundles was conducted to identify any common structures that NNETs or NETs employed with a significantly higher frequency. Similar to the results of functional analysis, there are similarities in the sub-structural categories of the two corpora despite the frequency differences. Table 7 shows the results of structural analysis of the key lexical bundles in the corpora.

TABLE 7
Frequency of Key Lexical Bundles by Structural Category

Structural category	NNET		NET	
	Freq.	Percent	Freq.	Percent
NP	32	10.3	11	8.7
(connector +) NP	27	8.7	8	6.3
NP + PP	5	1.6	3	2.4
VP	81	26	26	20.6
PP	15	4.8	4	3.2
(connector +) Cl	175	56.2	81	64.3
Others	8	2.6	4	3.2

As Table 7 indicates, the highest number of key lexical bundles in NNET and NET was from the *Cl (Clause)* category, i.e., 175 (56.2%) in NNET vs. 81 (64.3%) in NET. The second highest number of key lexical bundles was in the *VP (Verb Phrases)* category, followed by the *NP (Noun Phrase)* category. The frequent use of the *Cl* category may manifest consistent features of spoken discourse which included more clausal lexical bundles in spontaneous conversations (e.g., Biber & Barbieri, 2007; Biber & Conrad, 1999; Biber et al., 1999, 2004).

The combined percentages of the two most frequently employed lexical bundles, i.e., VP and clause-based key lexical bundles, were 82.2 in NNET and 84.9 in NET. The high percentage of VP and clause based lexical bundles was also reported in Biber and Conrad (1999). In addition, the distribution of key lexical bundles from both NNET and NET was similar in the sub-categories of the structural taxonomy.

Over-, Exclusive, and Under-use of Key Lexical Bundles

In order to examine key lexical bundle usages in contexts, the top-10 key lexical bundles were scrutinized and classified as over-used, exclusively-used, and under-used lexical bundles.

Over-use of Key Lexical Bundles in the NNET Corpus

The frequency analysis of the lexical bundles shows that the bundles of *you have to* and *I want you to*, both of which fall into the obligation and directive expressions category, show distinct frequency differences between the two corpora. The frequency of *you have to* is 383 in NNET and 95 in NET, and *I want you to*, 15 in NNET and 63 in NET. In order to examine bundles that can substitute *you have to* or *I want you to* in the corpus, the top-40 key lexical bundles were categorized into *obligation and directives* as in Table 8.

TABLE 8
Top-40 Key Lexical Bundles as Obligation and Directives Expressions in NNET and NET

NNET				NET			
Rank	Lexical Bundles	Freq.	Percent	Rank	Lexical Bundles	Freq.	Percent
1	you have to	383	35	2	you have to	95	8
6	you look at the	62	6	3	want you to do	74	7
7	listen to the	59	5	5	look at the	73	6
10	can you please	53	5	6	I want you	68	6
23	so you have to	59	5	7	I want you to	63	6
28	have to put	33	3	9	can you read	39	3
30	look at the	32	3	10	a look at	32	3
31	what don't you	31	3	16	you to do	28	2
32	you have to put	31	3	26	want you to do	19	2
				27	you had to	18	2
				31	what I want	18	2
				32	I want you to do	18	2
				37	ask each other	17	1

As is shown in Table 8, NNETs repeatedly used *you have to* and the lexical bundles including *you have to* in the corpus, e.g., *so you have to, have to put,*

you have to put. As a result, the lexical bundle *(so) you have to (put)* accounted for 46% of the obligation and directive expressions in the NNET corpus, while a more evenly distributed percentage of various types of the obligations and directives lexical bundles was witnessed in the NET corpus.

Another notable observation is that NETs used a structure that does not exist in the top-40 list of NNETs. For example, the lexical bundle of *want you to do* ranked high (ranked third) in NET but did not appear in NNET's list. In addition, the lexical bundles with *want you to do* frequently occurred in the NET's list such as *I want you, I want you to, want you to do, I want you to do*. Given the fact that the lexical bundle of *you have to* is considered more direct and less polite than the indirect request form, i.e., *want you to*, NNETs preferred a simpler and more direct structure over those denoting more politeness. NNETs' underuse of some of the lexical bundles could be explained by L1 intervention. In Korean, the first person pronoun *I* is not used in an imperative form or an obligation/directive expression (Sohn, 2013). Thus, imperatives using "I" such as *I want you to* may sound very awkward to NNETs when it is directly translated into Korean, which may have resulted in low frequency of the form in the corpus. Further, NNETs' high frequency of simpler forms may be partly due to the spontaneity of oral discourse. In other words, classroom teaching demands real-time production for NNETs (Shin, 2006), and thus using a simpler structure may lessen the real-time production burden of the teachers, which may in turn have resulted in the high frequency of the forms in the corpus. This might suggest that raising NNETs' awareness on useful lexical bundles that NETs employed in classroom situations may be necessary for non-native speakers in pre- or in-service teacher training programs.

A close examination of the NNETs' list suggested that lexical bundles such as *let's move* and *move on to* were often employed by NNETs to advance the teaching point by directing students to the next step. The lexical bundles of *let's move* and *move on to* were categorized into *topic introduction and focus* in the sub-category of *discourse-organizing expressions*. In particular, the word *move* was frequently observed in the NNET corpus, but lexical bundles

using *move* did not occur frequently in the NET corpus. Example (1) shows how *move* was employed in the NNET and the NET corpus.

(1) *let's move*
   In NNET (Frequency: 80)
   a. Okay? Alright! Okay, *let's move* on grammar section. (NNET1)
   b. Okay, *let's move* to next page. (NNET2)
   In NET (Frequency: 6)
   c. Okay, *let's move* along/on/away.

As shown in example (1), NNETs used *let's move* 80 times, but mainly with the preposition *on/to*, while NETs used it only 6 times, mainly with the preposition *along/on/away*. Some cases from NETs were followed by discourse markers such as "so" and "okay," and "please" which is added at the end of the sentence. The NETs used more variety of forms with the verb *move*, e.g., *we're gonna/we just/we'll/we/before we*. On the other hand, NNETs used two specific words of *let's* and *okay* mostly with *move on*. This different usages of *move to* reveal that NNETs relied much more on fixed expressions such as *let's move on/to* rather than using a larger variety of subjects and having to choose a different kind of temporal aspect.

Exclusive-use of Key Lexical Bundles in the NNET Corpus

The analysis of the key lexical bundles clearly shows NNETs' exclusive use of some of the lexical bundles. For example, lexical bundles such as *you can see*, *listen to the*, *can you please* are among NNETs' top-10 key lexical bundles, but have a zero frequency in NET. The frequency of these bundles was marked "0" because the minimum frequency was set as 5. The actual frequency of *you can see*, *listen to the*, and *can you please* are 4, 4, and 1, respectively in NET. Table 9 shows the exclusive use of key lexical bundles in NNET.

TABLE 9

Exclusive-use of Key Lexical Bundles in NNET

Key lexical bundles	NNET		NET	
	Examples	Freq.	Examples	Freq.
you can see	1. I wanna talk about "suggestion." There are several ways to make a suggestion in English. The first one is (as) *you can see* like "we should." Or you can also say, "you should." 2. Okay. So in this dialogue, *you can see* that there are comparative forms, right? Larger or better, or something like that. 3. (As) *you can see* here, how waitress greet here? "Hello, welcome to May kiss. My name is Kate. Are you ready to order?" 4. Is it black? Yeah. *You can see* it as is black. Okay, anyhow, this man chose the red one. Red one is better. Okay, and then, next item is? 5. Please take out your homework. I'm gonna check. (As) *you can see*, the homework, please check your answer there.	77	6. An outline is a list of the information you'll put in your essay. *You can see* an example of an outline on page 65. 7. That's the nice experience on tropical beach, swimming and *you can see* the bottom of the ocean, even though it is very deep. *You can see* fish very clearly. 8. Let's talk about the schedule. So, I hope *you can see* this. First of all, today's fifth, 31, right?	3
listen to the	(Please/let's) *listen to the* dialogue/exercise/differences.	59	*Listen to the* conversations/example.	2
can you please	*Can you please* read (number # questions)/look at/complete /pass/go (to the next page)?	53	*Can you please*?	1

As Table 9 indicates, NNETs frequently used *(as) you can see* when teachers were referring to a piece of information or a picture in the textbooks as shown in the examples 1, 2, and 3 of Table 9. As the examples 4 and 5 of

Table 9 show, *you can see* specifically means looking at something physically with your eyes by NNETs. However, NETs used as *you can see* only four times. Similar to NNETs, NETs used *you can see* to refer to a part or a picture in the textbook in examples 6 and 7. *"See"* means *understand* in example 8.

From these examples from Table 9, it can be assumed that NNETs more often referred to contents in the textbooks than NETs did by employing *(as) you can see*. NNETs often employed *(as) you can see* to ensure that students were looking at the right part in the textbooks. They might have used the lexical bundle since it can be smoothly translated into Korean and has a simple structure. As shown in Table 9, many uses of *you can see* were from *as you can see* in the NNET corpus, while NETs did not use this bundle. A look at the usage of *you can see* in conversations reveals that *you can see* had approximately 50 hits in BNC32 (Brezina, 2012)[3], but occurs only 1.1 times when normalized to 100,000. Thus, on average, *you can see* is not expected to appear as often as in conversational discourse from other native corpora.

In the second example of exclusive lexical bundle use by NNETs, NNETs frequently used *listen to the*. Given the fact that most of NNETs' classes employed multimedia materials such as CD or video clip to expose students to native speakers' speech, the frequency difference in *listen to the* may be due to having or not having a supplementary materials to listen to.

The lexical bundle of *can you please* was often used when NNETs asked students to work on a classroom task or when they wanted to draw students' attention to grammar points, as in Table 9. A sentence starting with *"can you"* implies an expression of politeness and *can you please* is even more polite. As discussed previously, NNETs used a bundle of *you have to* more frequently than NETs did, while NETs preferred *I want you to*. An assumption was made based on the different lexical bundle preference in that NNETs did not employ *I want you to* enough to show respect or politeness as

---

[3] BNC32 is a sub-corpus of BNC with a word count of 1 million. Brezina (2012) selected conversational texts from the BNC spoken corpus with a 10 million word count considering balanced samples such as gender, age, socioeconomic status, and a different regional dialect and compiled a sub-corpus called BNC32.

often as NETs did. However, NNETs expressed their politeness differently from NETs by employing *can you please* and *let's + verbs* (almost 600 occurrences) while NETs preferred to use imperative forms in similar functions. In terms of using *can you please*, NNETs seemed to be overly polite. The NNETs' overuse of the polite expression in class may be a reflection of pragmatic features of advanced learners, i.e., using politeness expressions much more than needed when requesting (Jalilifa, 2009).

Under-use of Key Lexical Bundles in the NNET Corpus

Under-use of key lexical bundles in the NNET corpus was examined by calculating negative keyness value of the bundles. As Scott (2012) noted, "a lexical bundle which has negative keyness appears less often than would be expected by chance in comparison with the reference corpus." From the key lexical bundle list of the NNET corpus, eight lexical bundles were found to have negative keyness value, which clearly shows NNETs' underuse of those bundles compared to the corresponding bundles of NETs. Table 10 presents key lexical bundles with negative keyness in the NNET corpus.

TABLE 10
Key Lexical Bundles with Negative Keyness in NNET

Lexical Bundle	NNET		NET		Keyness
	Freq.	Percent	Freq.	Percent	
and then you	6		29	3	-15.7059
what did you	31	3	73	6	-16.1105
can you read	6		39	3	-25.9747
I want you to	15	1	63	6	-30.214
do you think	68	6	155	14	-32.0916
want you to	18	2	74	7	-34.7825
I want you	15	1	68	6	-34.9218
how would you	11	1	63	6	-38.6952

A close examination of the corpus through the concordance analysis rendered qualitative differences of three lexical bundles *what did you, do you think* and *how would you* in the NNET and NET corpus. First of all, NETs frequently asked questions using *what did you*. In order to examine how the lexical bundle *what did you* was used in the context, the KWIC search for the bundle was conducted. Figures 2 and 3 show the results of the KWIC search for *what did you* in the NNET and the NET corpus, respectively.

N	Concordance
1	you. Let's hear your example, please. What did you buy? *T: Old golf racket? Golf club?
2	So, they have the questions at the bottom. What did you buy? Why did you buy it? Do
3	here, Mr. New Jacket. Alright. Tell me what did you buy, tell us. Okay. Okay. Okay. It's
4	never. Okay. So, piano, golf clubs. What did you buy? *T: Ah, the vacuum. Robot
5	. Very good example. *T : Mr. Park. Tell me what did you buy that you regret? Sorry? Ah,
6	Go for WJ. are you still with us? Number 2. What did you choose? No, you need to stay
7	you still have it? If yes, where is it? If not, what did you do with it? Would you ever buy
8	do anything for Halloween? By, yes, what did you do for Halloween? Okay, alright, so
9	that's like a philosophical, oh, I'm nothing. What did you do? You went to Paju? Okay.
10	That's easy. We talk about that all the time. What did you do last weekend? What will you
11	are not okay. I can't take. It was so-so. What did you do over the weekend? It was
12	. How was your weekend? It was terrific. What did you do on Monday? Grammar, okay.
13	you do over the weekend? It was not good. What did you do last weekend? It was good.
14	are you going to do this weekend? So, what did you do last weekend? Sleep and eat?
15	answers. Can't take it. Alright. Next one. So what did you do this weekend? It was pretty
16	I see you what do I ask you. JK? Yes. What did you do this weekend? JK, what did you
17	item? I told you on Monday, we ask what did you do this weekend. Tuesday,
18	. Wednesday what do I ask? What did you do last weekend? Okay. WH. WH you
19	talk to you today. Yeah, there's no passing. What did you do this past weekend? Ah. Ah.
20	. What did you do this weekend? JK, what did you do this weekend? What if it's on, no

FIGURE 2
Concordance Lines for *what did you* in NET

N	Concordance
1	can ask, you know, about preference. About what did you think of, what do you think of,
2	No? How about HJ? In subsection B. what did you write down? Right. So they
3	good. how about sub topic second one B. What did you write down? LJ. What did you
4	and they already give us title Baptism? What did you write down? For that? JH what
5	nice night view? Like noisy. 4 story building? What did you say? Like crowded, lot of
6	Sandra? Clothing? About her Clothing? What did you write? White dress. Anything else?
7	BC did you listen to the conclusion? What did you write down for the conclusion
8	girl as her cousin right? How about DB? what did you write down for Sandra? Music? Right.
9	, event? Going to the college, what else? What did you say, YJ? Yeah, get married.
10	Okay, can I move to the grammar? What did you learn about the grammar in the handout?
11	(L1 T). Okay, let's go to grammar. What did you learn last week's grammar handout? Did
12	right? Violet purple. Right? How about, HJ? What did you write about Katu's preference?
13	you fill out of all the blank? Blank? HS, what did you write about Sandra's wedding
14	your turn? JY's turn. 22. Okay. 1,2,3. Huh? What did you say? Say that. No. congested.
15	Which movie do you wanna see? Huh? What did you say? What movie do you want
16	so she chose a classical guitar and flute. IY, what did you write about Sandra's
17	For the third part at Baptism how about JH? What did you write down to the third part.
18	What did you write down? For that? JH what did you write down? For the third part
19	white or pink. Pink flowers. Right. JS, what did you listen about Katu's wedding
20	grammar handout? Did you learn? What did you learn? You learned passive voice. You

FIGURE 3
Concordance Lines for *what did you* in NNET

As Figure 2 shows, NETs mainly asked casual questions to warm up classes or elicit students to volunteer by suggesting additional points towards students' grades. The concordance line search shows that a total of 73 lines produced examples of *what did you* from NETs, and in many of them, teachers were asking their students what they did over the weekend. NNETs, however, as shown in Figure 3, did not ask their students those questions as often as NETs did, and they mostly used *what did you* to ask students to share an opinion, or to clarify answers related to topics or questions from the textbooks.

The examination of the NNET corpus revealed that NNETs used almost double the frequency of *do you think* than NETs (155 in NET vs. 68 in NNET). As shown in the following example (2), *do you think* in the NET corpus mainly used as an insertion in wh-questions followed by a clause. Sentences in example (2) were used to ask students' opinions of topic-related items or expressions when beginning a new unit.

(2) From the concordance lines of *do you think* in NET
  a. When/how/what/why *do you think* that S + V
    When do you think a practical joke crosses the line?
    How do you think these things will change in the future?
    What do you think it means?
  b. *Do you think* S + V
    Do you think globalization is a good thing for the world?
    Do you think it's a good answer for this?

On the other hand, NNETs frequently employ the structure of *do you think* followed by a clause as seen in example (3a) and (3b).

(3) From the concordance lines *do you think* in NNET
  a. *Do you think* S + V
    Do you think I have to speak slowly?
    Do you think you agree or disagree?
  b. How/what/why/who *do you think* that S + V
    What do you think the man and woman are talking about?
    Why do you think A is the answer?
  c. More formulaic and simple forms of *do you think*
    What do you think (of/about)?
    What do you think of (something/anyone: David, his or her opinions)?

NNETs might have expected students to answer simply, e.g., with "yes" or "no," or with short answers. Alternatively, NNETs may have found it easier

to use this structure than to attempt to insert *do you think* in wh-questions. The lower frequency of *do you think* in wh-questions may enable us to infer that NNETs prefer a simpler structure and have difficulties using wh-questions when inserting *do you think* as a filler.

The observation of the NNET corpus obviously witnessed examples of wh-questions with *do you think*, as in example (3c). The wh-questions in the NNET corpus, however, had a shorter length of phrases. They often used "what do you think?" as a single phrase to ask students' opinions and "what do you think of/about" as fixed expressions. The NNETs added a name of a person or personal pronouns at the end of "what do you think of" instead of using long clauses.

Another interesting feature in NNET's underuse of lexical bundles is the use of *how would you*, as in example (4).

(4) *how would you* in NNET
   How would you say (6)/ start your dialogue (2) /respond: from NNET4
   How would you like to pay for (them/loafers) (2): from NNET1

NNETs employed *how would you* only 11 times, while NETs, 63 times. In addition, NNETs' use of *how would you* was limited to two NNETs, i.e., NNET1 and NNET4. A close examination of the data showed that NNET1's sentences were influenced by the textbook scripts.

On the contrary, NETs had a much higher number of 63 entries. All the NETs, except for NET4, employed *how would you* in their classes. Example (5) shows various sentences starting with *how would you*.

(5) *how would you* in NET
   How would you answer that/change it/define humor/describe this (it/number #/that)/explain that/it to me/make it better/say that verbally/say formally/say the answer?

The underuse of *how would you* indicates that NNETs may have

difficulties using hypothetical statements or questions using "would" independently from *if*-clauses. NNETs did not seem to be familiar with using *would*-clauses without *if*-clauses. Developing a question using "would" was expected to be more difficult for NNETs than making a statement using "would."

In order to examine how NNETs make statements using "would," "would" was used as a search word in the KWIC search. Example (6) shows how "would" is used in the context. From the concordance lines of the KWIC search results, NNETs seemed to know how to use the *would*-clauses without *if*-clauses in a statement form. However, NNETs often used the *would*-clause in tight connection to the *if*-clause, as seen in example (6c) and (6d).

(6) A search word of *"would"* in NNET
   a. I think the vocabulary section *would* be more difficult one to do. No? (NNET1)
   b. Okay, I *would* say I'm late this morning. (NNET4)
   c. I was wondering <u>if</u> it *would* be possible to. (NNET2)
   d. <u>If</u> I were you, I *would* go home and check again. (NNET2)

The under-use of *would*-clauses without an *if*-clause is in accordance with the results of previous studies (Frazier, 2003). As Frazier (2003) suggested, second language learners' under-use of *would*-clause without an *if*-clause may be due to the influence of ESL/EFL textbooks presenting *would*-clauses together with *if*-clauses. NNETs' under-use of *would*-clauses without an *if*-clause shows the need for greater awareness on part of NNETs on the different uses *if*-clauses and *would*-clauses when making hypothetical statements by NNETs and NETs.

CONCLUSION

The study investigated the use of lexical bundles in the non-native and native English teacher talk corpus. In particular, key lexical bundles of

NNETs and NETs were extracted and compared to each other in order to examine closely the functional and structural features of lexical bundle uses. The results of the functional analysis indicated that key lexical bundles as *stance expressions* were observed the most, i.e., more than 50 percent of the key lexical bundles, and as *referential expressions*, the second most. The structural analysis showed that the clause-based key lexical bundles were most frequently observed, occupying about 56% in NNET and 65% in NET. Among phrase-based key lexical bundles, verb-phrase based key lexical bundles accounted for the greatest number, which had similar percentage of lexical bundles in conversations of previous studies. In addition, the key lexical bundles were discussed as over-used, exclusively-used, and under-used lexical bundles. One of the most notable differences between the NNET and NET corpus was found in over-used lexical bundles: *you have to* vs. *want you to do* in NNET and NET, and *let's move* and *move on to* in the NNET corpus. NNETs seemed to overuse a limited number of lexical bundles as they were confident using them. The NNETs know how to use the counterpart lexical bundles that the NETs overly used, but they used them at much lower frequency than NETs did. The NNETs in the study seemed to rely more on lexical bundles that they felt confident and comfortable in using. As Ellis (2012) stated, on "phrasal teddy bear", L2 speakers tended to use those words or phrases that they felt confident about. In this study, NNETs as teachers as well as learners clearly showed that they depend much more heavily on those lexical bundles that required less cognitive load. In other words, over-used lexical bundles were already acquired in NNETs and could be retrieved right away without hesitation. Further, the exclusively-used lexical bundles were *you can see, listen to the,* and *can you please,* and under-used lexical bundles, *what did you, do you think* and *how would you*. Interestingly enough, a close examination of the exclusive- and under-use of lexical bundles revealed that qualitative differences exist between NNET and NET. The results of the study indicate that novice non-native teachers preparing for teaching EFL classes might need to acquire more lexical bundles to base their classroom talks on. Experienced non-native teachers

need to be drawn to over- and under-use of lexical bundles from authentic teacher talk in EFL classes. In particular, NNETs' under-use of a certain limited number of lexical bundles indicated learners' difficulties with phraseological patterns (Granger, 2012). While a native corpus could show typical usages of lexical bundles, a comparative study based on a learner corpus and a native corpus could highlight the spots to draw the learners' attention to. The EFL teacher corpus was designed specifically to instruct EFL non-native teachers and would be a useful learner corpus for non-native teachers to reflect on their own language practices and prioritize phraseological patterns.

Different lexical bundle preferences suggest specific lexical bundles that NNETs need to complement in order to speak more fluently in teaching EFL classes. The lexical bundles that NETs preferably employed are those that NNETs know how to use, as those lexical bundles were witnessed in NNET, but with lower frequency rate. Therefore, awareness raise on these lexical bundles could enhance NNETs' proficiency. Even though a limited number of lexical bundles was extracted for detailed examination, those differences would have been neglected without a corpus-based study providing frequency information, since NNETs did not misuse those lexical bundles, but rather, overused or underused them in the current study. As Krishnamurthy (2002) claimed, it would be very difficult for learners to notice a lexical bundle unless they were drawn to it. Therefore, awareness-raising on over- and under-uses of lexical bundles could help non-native teachers identify exactly what aspects of their talk needed to be polished.

THE AUTHORS

Ye-Eun Kwon is a teaching professor in the Institute of Educational Development at Kunsan National University. She received her Ph.D. in English Education at Ewha Womans University and M.A. in Professional Education at SUNY Potsdam. This article is based on a part of her doctoral

dissertation. Her main research interests include applied corpus linguistics and teacher education. She is the first author for this article.

Institute of Educational Development
Kunsan National University
558 Daehakno, Gunsan, Jeobuk 573-701, Korea
Phone: +82 634694998
Email: yeeun78@daum.net

Eun-Joo Lee is a professor in the Department of English Education at Ewha Womans University. She received her Ph.D. in applied linguistics from UCLA and M.A. in ESL from the University of Hawaii at Manoa. Her main research interests include applied corpus linguistics, lexical processing and acquisition and second language acquisition. She is the corresponding author for this article.

Department of English Education
Ewha Womans University
52 Ewhayeodae-gil, Seodaemun-gu, Seoul 120-750, Korea
Phone: +82 232772647
Email: eunlee@ewha.ac.kr

REFERENCES

Baker, P. (2009). The question is, how cruel is it? Keywords, fox hunting and the house of commons. In D. Archer (Ed.), *What's in a word-list? Investigating word frequency and keyword extraction* (pp. 125-136). Farnham: Ashgate.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263-286.

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora: Studies in honor of Stig Johansson* (pp. 181-190).

Amsterdam: Rodopi.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English*. London: Longman.

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research, 10*(3), 245-261.

Brezina, V. (2012). BNC 32. Retrieved from the World Wide Web: http://www.lknol.com/BNC32/BNC.html.

Conklin, K., & Schmitt, N. (2008). Formulaic Sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics, 29*(1), 72-89.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397-423.

Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora, 3*(1), 43-57.

DeCock, S. (2004). Preferred sequences of words in NS and NNS speech. *Belgian Journal of English Language and Literature*, *2*, 225-246.

Doyle, P., & Hong, H. (2009). *Compiling a multimodal corpus database of educational discourse in Singapore schools* (Report No. CRPP Project CRP 20/04 JD / CRP 7/05 AL / CRP 13/05 AL). Singapore: Centre for Research in Pedagogy and Practice, National Institute of Education.

Ellis, N. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics, 32*(1), 17-44.

Frazier, S. (2003). A corpus analysis of would-clauses without adjacent if-clauses. *TESOL Quarterly, 37*(3), 443-466.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 145-160). Oxford: Oxford University Press.

Granger, S. (2012, October). *Towards a phraseological model of lexical proficiency: Evidence from learner corpus research.* Paper presented

at the 2012 International Conference of the Applied Linguistic Associations of Korea. Seoul, Korea.

Granger, S., & Paquot, M. (2010). Customising a general EAP dictionary to meet learner needs. In S. Granger & M. Paquot (Eds.), *eLexicography in the 21st Century: New challenges, new applications* (pp. 87-96). Louvain-la-Neuve: Presses universitaires de Louvain.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics, 19*(1), 24-44.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*(1), 4-21.

Jalilifar, F. (2009). Request strategies: Cross-sectional study of Iranian EFL learners and Australian foreign speakers. *Canadian Center of Science and Education, 2*(1), 46-61.

Kemppanen, H. (2004). Keywords and ideology in translated history texts: A corpus-based analysis. *Across Languages and Culture, 5*(1), 89-106.

Krishnamurthy, R. (2002). Language as chunks, not words. *JALT 2002 Proceedings* (288-294). Retrieved from the World Wide Web: http://jalt-p publications.org/archive/proceedings/2002/288.pdf.

Ku, P., & Yang, A. (1999). An analysis of key words in tourism English. *Proceedings of the 1999 English for Specific Purposes Conference* (pp. 232-241). Taiwan: Ming Chuan University.

Mauranen, A. (2006). Speaking the discipline. In K. Hyland & M. Bondi (Eds.), *Academic discourse across disciplines* (pp. 271-294). Bern: Peter Lang.

McEnery, T. (2009). Keywords and moral panics: Mary Whitehouse and media censorship. In D. Archer (Ed.), *What's in a word-list?: Investigating word frequency and keyword extraction* (pp. 93-124). Farnham: Ashgate.

Scott, M. (2012). *WordSmith tools 6.0.* Oxford: Oxford University Press.

Seal, C., Charteris-Black, J., & Ziebland, S. (2006). Gender, cancer experience and internet use: A comparative keyword analysis of interviews and online cancer support groups. *Social Science and Medicine, 62*(10), 257-259.

Shin, D. K. (2006). *A collocation inventory for beginners* (Unpublished doctoral dissertation). Victoria University, Wellington.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487-512.

Sohn, H. (2013). *Korean: Descriptive grammars.* Seoul: Korea University Press.

Tribble, C. (2011). Revisiting apprentice texts: Using lexical bundles to investigate expert and apprentice performances in academic writing. In F. Meuniers, S. De Cock, G. Gilquin, & M. Paquot (Eds.), *A taste for corpora: In honor of Sylviane Granger* (pp. 85-108). Amsterdam: John Benjamins.

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 155–172). Amsterdam: John Benjamins.

Wood, D. (2006). Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency. *Canadian Modern Language Review, 63*(1), 13-33.

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

Wray, A., & Fitzpatrick, T. (2008). Why can't you just leave it alone? Deviations from memorized language as a gauge of nativelike competence. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 123-147). Amsterdam: John Benjamins.