Lexical Bundles in Economic Research Articles by Native and Non-native Speakers of English

Bit-Na Choi

1. Introduction

When asked what is the most remarkable event in the linguistic history in the twentieth century, one would suggest the phenomenal spread of English across the globe. With the major historic change over the centuries that the world in the past fell under the British and North American rule, internationalization caused the spread of English and its increased use for communicative functions by non-native speakers (Hoffman, 2000). Given its status as *lingua franca* (Hoffman, 2000), the presence of and need for English has been far-reaching in all areas, particularly in academia, where conventionalized high-proficiency English is required. One cannot overlook the fact that a growing body of research is conducted and that a dramatically increasing amount of academic texts, as a consequence, are produced by non-native speakers of English (Römer, 2009). As the nativeness of non-native writers has emerged as a significant issue, a multitude of studies have explored the elements of well-written prose and the ideal way to teach them English in academic contexts (Salazar, 2011). In the center of it lies the issue of frequent word combinations whose use has been regarded as a marker of proficient language use of a particular register including academic writing (Cortes, 2004). For instance, Altenberg (1998) found in a search of the London-Lund Corpus that approximately 80% of words in the corpus constituted part of recurrent word combinations.

Recurrent word combinations have been adopted in various forms by dif-

ferent scholars. Some of them include *clusters* (Hyland, 2008a), *recurrent word combinations* (Altenberg, 1998; Lindquist, 2009), *n-grams* (Stubbs, 2007a), and *lexical bundles* (Biber & Barbieri, 2007). These terms refer to continuous word strings retrieved by taking a corpus-driven approach with specified frequency and distribution criteria. The retrieved recurrent sequences are fixed multi-word units with conventional pragmatic and/or discourse functions, used and recognized by the speakers of a language within certain contexts (Chen & Baker, 2010). Throughout the study, *lexical bundle* is adopted as the primary term to refer to recurrent word sequences that occur in academic register, as has been used by Biber in his series of studies on which the theoretical framework of the current study is based.

2. Literature Review & Purpose of Study

Since the beginning of the last century, the study on lexical bundles has caught the eye of numerous researchers. In a series of lexical bundles research by Biber and colleagues (Biber & Barbieri, 2007; Biber, Conrad, & Cortes, 2004; Biber, 2006), they found that conversation and academic prose displayed different use of lexical bundles. Cortes (2004) explored the difference in the use of lexical bundles in academic prose by published authors and by students at three different levels in history and biology. It was found that students rarely used the lexical bundles which were used by published authors. If they did, they used it in a different way. Hyland (2008b) examined the structures and functions of research articles, doctoral dissertations and master's theses in four disciplines. He found that there existed disciplinary variations in the use of lexical bundles across disciplines. In another research he conducted on published academic prose and postgraduate writing, Hyland (2008a) observed that postgraduate students were more likely to adopt formulaic expressions than native scholars.

In 2010, Chen and Baker followed the line of research and investigated different uses of lexical bundles in academic writing. They targeted three groups — published authors, native student writers and non-native student writers. A wider range of use was found in writings of published authors than that of

students and the two student authors displayed some distinctive uses of their own in lexical bundles. Ädel and Erman (2012) expanded the scope of research to the structures and functions of lexical bundles in writings of native speakers and advanced learners. They reported that a larger number of and a more varied lexical bundles were employed by native speakers, which was similar to the phraseological research tradition in SLA. Lee (2012) conducted a study on the use of lexical bundles in academic writing by native and non-native speakers and found from some structural differences that non-native writers displayed the overuse and the underuse of certain expressions. Salazar (2011), in his doctoral dissertation, demonstrated the use of lexical bundles that non-native scholars displayed overuse, unnecessary repetition, lack of variation, and in particular a limited use of participant-oriented bundles.

As presented, a majority of previous researches on lexical bundles in academic register have mainly focused on the disciplines of linguistics and life sciences. Linguistics has been selected due to the texts availability by the researchers, and life sciences for the difficulty faced by many non-native scientists with its high possibility for application. While following the framework, this paper moves the realm to economics, which lies on the border of arts and sciences and whose research can present non-native economists a general idea of formulaic expressions in research articles.

Thus, this study aims to answer to the following questions:

1) What are the most frequent four-word lexical bundles in published economic articles of native and non-native speakers of English? How are these lexical bundles structurally classified?
2) Are there any significant differences in the structures of these expressions between native and non-native writers?

3. Material and Method

3.1 The corpus material

Economics was selected as the target academic discipline. Texts were collected from the journals by two economic institutes that represent the United States and Korea. The Brookings Institution is an American think tank based

Table 1. Material for the native and non-native-speaker subcorpora

	Native speaker corpus	Non-native speaker corpus
Texts	11	13
Words	165,566	97,087
Average text length	15,051	7,468

in the United States and conducts research and education in a variety of areas, primarily in economics, and others as well, including the social sciences, foreign policy, and global economy and development (http://en.wikipedia.org/wiki/Brookings_Institution 2013). The KIEP, which stands for the Korea Institute for International Economic Policy, is a think tank under the affiliation of the South Korean government. The research focuses on the relations between Korea and the international economy (http://en.wikipedia.org/wiki/Korea_Institute_for_International_Economic_Policy). Both institutions have similarities in that they produce economic periodicals as the leading think tanks for each government. Authors of the periodicals are nationwide renowned researchers. In this sense, the subcorpora were considered comparable to each other.

The database includes 11 papers from the journals published by the Brookings Institution and 13 from the KIEP, so as to represent one native and the other non-native authors of English. Recent articles ranging from 2010-2013 were selected for generalization of the results. The articles from the Brookings Institution contained 165,566 words in total while those from the KIEP had 97,087 words, amounting to 262,653 words combined. Table 1 gives an overview of the data used for the subcorpora.

3.2 Analytical steps

As for target bundles, four-word lexical bundles were considered for the study. It is known to be "the most researched length for writing studies, probably because the number of four-word bundles is often within a manageable size for manual categorization and concordance checks" (Chen & Baker, 2010, p.32). KfNgram (Fletcher, 2010) was used to retrieve corpus query.

In respect to how to create a list of lexical bundles, two key criteria have been identified in the literature. The cut-off frequency determines the num-

ber of lexical bundles to be included in the analysis. The frequency adopted for lexical bundle identification has varied from 10 to 40 per million words (henceforth, PMWs); Biber et al. (1999) 10 PMWs; Biber and Conrad (1999) and Hyland (2008) 20 PMWs; and Biber et al. (2004) 40 PMWs. Second is the dispersion criterion, the requirement for the lexical sequences to occur across texts. It usually varies from at least 3-5 texts, or 10% of texts. Since this research is a pilot study utilizing considerably less articles, neither frequency nor dispersion criteria were considered to avoid overgeneralizing the results. However, the bundles resulted from the study appeared at least 3 times in at least one subcorpus.

By referring to the previous studies, the retrieved bundles were checked manually for exclusion; Content bundles including proper nouns, topic-related terms and expressions unique to the discipline were excluded. However, terms and expressions for research in general remained.

4. Analysis & Results

4.1 Structural taxonomy

The structural classification of lexical bundles followed that in the Longman Grammar of Spoken and Written English (henceforth, LGSWE) (Biber et al., 1999), the generally and widely adopted taxonomy in the studies on recurrent word sequences (Cortes, 2004; Hyland, 2008a, 2008b; Chen & Baker, 2010; Lee, 2012).

Three broad structural categories were identified: "NP-based," "PP-based," and "VP-based." Another category created was "others," for those that did not belong to any of them, such as *as well as the*. NP-based bundles consist of noun phrase components, usually ending with the beginning of a post-modifier (e.g. *the null hypothesis of, an important role in)*. PP-based bundles involve prepositional phrase components with embedded modifiers such as *of rule of law, on the basis of*. VP-based bundles include any phrases which contain verb phrase components. The "others" category is for the rest, which do not belong to any of the aforementioned categories.

Table 2. Distribution of lexical bundles across the structural categories from LGSWE (Biber et al., 1999; Chen & Baker, 2010)

Category	Pattern	Example
NP-based (noun phrase with post-modifier fragment)	NP with "of" phrase fragment NP without "of" phrase fragment	*the null hypothesis of* *an increasing number of* *an important role in* *general and bilateral exemptions*
PP-based (preposition + noun phrase fragment)	PP with "of" phrase fragment PP without "of" phrase fragment	*on the basis of* *of rule of law* *in the sense that* *in appendix i except*
VP-based	1) Copula *be* + NP/AdjP 2) VP with active verb 3) Anticipatory *it* + VP/adjP + complement-clause) 4) Passive verb + PP fragment 5) (VP +) *that*-clause fragment 6) (V/Adj +) *to*-clause fragment 7) Others	*is equal to the* *does not seem to* *it is expected that* *it seems likely that* *are explained by the* *as seen in the* *we assume that the* *sense that it is* *this study is to* *more likely to form* *if we substitute equation*
Others		*as well as the* *a positive and significant*

4.2 Distribution of lexical bundles across the structural categories in the native and the non-native-speaker subcorpora

In table 3, it can be seen that the uses of NP-based bundles and VPs manifest marked differences in its tendency. Whereas both groups use PP-based bundles to the same extent, the NS show the highest use in NP-based bundles, 14% more than the non-native writers in raw difference, which accounts for a 35% actual difference. On the other hand, the NNS exhibit just the opposite trend; they display a noticeably higher use in VP bundles than the native ones – 16.5% in raw difference and 51.5% actual difference. Along with them, another huge difference lies in the "others" category. The NS had 25 types of bundles while the other group had one, resulting in 62.5% of difference. Considering the fact that the number of words in the non-native corpus is a little less than the twice of that of the NS, it is a large difference.

Table 3. Proportional distribution of lexical bundles across the structural categories in the native and the non-native-speaker subcorpora

Pattern		Native subcorpus		Non-native subcorpus	
		Types	%	Types	%
NP-based	NP with "of" phrase fragment	137	21	45	17
	NP without "of" phrase fragment	120	19	23	9
NP in total		257	40	68	26
PP-based	PP with "of" phrase fragment	75	11	32	12
	PP without "of" phrase fragment	85	13	31	12
PP in total		160	24	63	24
VP-based	1) Copula *be* + NP/AdjP	57	8.5	23	8.8
	2) VP with active verb	53	8	22	8.4
	3) Anticipatory *it* + VP/adjP + complement-clause)	6	1	8	3
	4) Passive verb + PP fragment	50	7.7	43	16.4
	5) (VP +) *that*-clause fragment	16	2.6	11	4.2
	6) (V/Adj +) *to*-clause fragment	23	3.5	19	7.3
	7) Others	5	0.7	1	0.4
VP in total		210	32	127	48.5
Others		24	4	1	1.5
Total		651	100	262	100

The result that the NS prefer NP-based bundles to VP-based bundles in comparison of NNS lends support to the finding of Chen & Baker (2010), who reported that the NS showed a higher use of NP-based bundles than the NNS while the NNS employed a lot more VP-based bundles when compared to native expert authors.

4.3 VP-based bundles

When it comes to VP-based bundles, the NNS manifest a far higher use compared to the NS. Of their uses, a salient distinction is made by epistemic markers.

As revealed in table 4, the NS employ a higher number of and a wider variety of epistemic markers. The NNS use them 27 times in all but the NS 35 times. While the NNS show a limited use clinging to 6 types of epistemic markers (*likely, can, seem, assume/assumed, thought, could*), the NS take an

Table 4. The use of epistemic markers in the NS and the NNS corpora

	NS	NNS
likely	12	10
can	5	7
seem	2	6
assume/assumed	1	2
thought	1	1
could	2	1
would	8	0
might	1	0
may	1	0
appear	1	0
perhaps	1	0
Total	11 types/35 tokens	6 types/27 tokens

Table 5. The use of anticipatory *it* + VP/adjP + complement-clause verb bundle in the NS and the NNS corpora

NS	NNS
It is important to	It is clear that
It is difficult to	It is not clear
It is possible that	It is obvious that
It is not clear	It is important to
It is reassuring that	It seems likely that
It seems reasonable to	It is expected that
	It could be hypothesized

advantage of variation in expressing the degree of certainties of their claims. This result converges with the findings of previous researchers (Chen & Baker, 2010; Ädel & Erman, 2012; Lee, 2012) that the NS tend to employ a larger number of and a wider range of epistemic markers, or hedges.

Another discrepancy can be found in the use of the anticipatory *it* + VP/adjP + complement-clause verb frame. In examining the bundles in table 5, the matter of certainty again arises as an issue. Bundles adopted by the NNS such as *clear, obvious, important, likely, not clear* all point to certainty/uncertainty. However, two distinctive bundles from the NS corpus are nonexistent in the NNS data, *it is difficult to* and *it is important to.* By looking at the concordances lists can one know that they function as hedges that weaken the certainties of certain statements.

(1) **It is difficult to** *reach definitive conclusions about its impact.*

(2) *Given this,* **it is difficult to** *estimate a target level of leverage for any given household with confidence. However, one reasonable benchmark for such a target might be the level of mortgage leverage the household had in the precrisis period.*

(3) F*or example, to the extent that NCLB increased the expectations for academic achievement in states without prior school accountability policies,* **it is possible that** *teachers simultaneously chose to benchmark the behavioral engagement of their students with school against a more lax standard. If this is true, our estimate of the impact on behavioral engagement is biased upward.*

(4) **It is possible that** *NCLB led to a level shift in student achievement, which would be manifest as a shift in the intercept after NCLB. It is also possible that NCLB changed the rate of achievement growth,*

As seen in the examples, the *difficult* bundle is adopted when the author is being cautious in presenting a claim or an argument. It is to reduce the risk that might result from drawing a definitive conclusion from the results. In (3), the *possible* bundle is employed to note that the author's analysis may be skewed. From (4), it is confirmed that the author uses the *possible* bundle when introducing more than one potential in developing analysis of results. In other words, the bundles function as hedges as seen in nowhere in the texts of the NNS.

Another way in which the two groups differ is in the use of "*we* + verb" bundles. The bundles are either used to report the procedure or to present research results and discussion. Of the bundles that play the latter role, one significant distinction is made between the NS and the NNS groups. In introducing results and discussion, the NS tend to employ the *we* + verb form such as *we find that the, we do not find, we show that the.* However, the NNS do not adopt the form but instead *we* + modal + lexical verb as in *we can derive the, we can see that, we assume that the, we can argue that.* From this, one can conjecture that NNS have reluctance in using the form *we* + verb to signal the ownership of the results and conclusions being made. This differentiation can be attributed to the traditional concept of academic prose as being objective and highly impersonal, a view that has been implanted into non-native authors by numerous writing manuals and style guides (Harwood, 2005a; Salazar, 2011).

It has been examined and found as such in the previous literatures (Salazar, 2011).

Along with these, particular attention needs to be made in how the two groups differently use modifying bundles. In using modifiers before certain words, the NS show much more variation than the NNS. *Likely* is an example as such. While the NNS use only three types in *likely* (*are likely, more likely, seems likely*), the bundle types the NS employ are six (*more likely, most likely, most unlikely, less likely, somewhat likely, somewhat or very likely*).

4.4 NP-based bundles

A 14% raw difference exists between the NS and the NNS data regarding NP-based bundles. This can firstly be attributed to the NS' tendency to use bundles that include certain words to a greater extent and with higher variation than the NNS. *Effect* and *change* are the examples.

Table 6. Types of bundles that include *effect* and *change* in the NS and the NNS data

Effect		Change	
NS (20 types)	NNS (2)	NS (9)	NNS (1)
The effect of the	That the effects of	The change in the	To the change of
The effects of the	A positive effect on	A change in the	
The effects of the		The changes in the	
The effects of a		And changes in the	
That the effects of		And the change in	
An effect of about		Or a change in	
An effect of those		Changes in the relative	
The likely effects of		Changes over time in	
The initial effects of		The joint change in	
Effect of the recent			
Effects of changes in			
Size of the effects			
The overall effect of			
Cumulative effect of all			
Cumulative effect of the			
And calculated effects of			
The estimated effects of			
An effect on the			
Little effect on the			
A large effect on			
Large effect on the			

It is displayed in table 6 that the NS use 20 types of bundles which involve *effect* and 9 types with *change* and the NNS use only 2 and 1 respective types. Counting the fact that the NS corpus is 1.6 times larger than the NNS data, the NS use the *effect* bundles 32 times more than the NNS and the *change* bundles 14.4 times higher.

The use of quantifying bundles is another factor which makes a notable distinction between the NS and the counterpart. The NS, unlike the NNS, employ a vast number of and a wide range of quantifiers to modify noun phrases. The examples can be seen in the following:

> (5) *a great deal of*
> *the vast majority of*
> *a large majority of*
> *a small portion of*
> *a larger fraction of*
> *a small share of*
> *a share of all*
> *a wider range of*
> *a closer look at*
> *statistically significant impact on*
> *a significant impact on*
> *a substantial increase in*
> *a permanent increase in*

Added to this, another notable difference lies in the bundles that direct readers along the text. In the NS corpus, a large number of bundles are used to perform the function whereas the NNS employ only two. Some of the examples from the NS corpus are given below:

> (6) *The lower end of*
> *The third panel of*
> *Panel of the figure*
> *Panel of the table*
> *The bottom panel of*
> *Bottom left panels of*
> *Bottom right panel of*
> *The first two columns*

The top quintile of
The top quartile of
The right-hand side of
Right-hand side of equation
Appendix for further details
Baseline specification described in

4.5 PP-based bundles

One striking feature that differentiates the NS and the NNS in the use of PP-based bundles is in the presence of "lexical teddy bears" (Hasselgren, 1994). It refers to the NNS' reliance on a term or an expression which is familiar, by "choosing words and phrases closely resembling their first language or those learnt early or widely used" (Hasselgren, 1994, p.237). It has been suggested in numerous studies that learners tend to hold on to lexical teddy bears in academic writing (Ädel & Erman, 2012; Hasselgren, 1994). Table 6 describes the use of lexical teddy bears by comparing the proportion of use of the NNS and the NS.

The bundles in table 6 mark at least 2 times of difference in use in the NS and the NNS subcorpora. The following are some of the concordances lists of *in other words the,* which made nearly 4 times of difference in frequency between the NS and the NNS.

(7) The results in this panel indicate that just turning off the effect of UI extensions on labor force exit reduces unemployment by more than half as much as did turning off both UI effects in the top

Table 6. The proportion of use of lexical bundles in the NS and the NNS subcorpora

Type	Proportion of Use in the NS Corpus	Proportion of Use in the NNS Corpus
on the other hand	11	40
as a percentage of	0	22.5
the null hypothesis of	0	21
in other words the	1	16
of this paper is	4	13
the rest of the	4	13
in the form of	8	18
at the same time	8	16

panel.31. *In other words, the* majority of the effect of UI extensions on overall unemployment and on long-term unemployment operates through the labor force exit channel, by keeping people in the labor force who would otherwise have exited, rather than through reduced reemployment rates. [NS]

(8) This is how confidence of investors about stock prices in the future affects the unemployment, *in other words, the* level of economic activity. [NNS]

A key difference exhibited in concordances lists in the use of *in other words the* is that the NS mostly use the bundle at the beginning of a sentence, paraphrasing the preceding sentence as to help readers grasp the message. In contrast, the NNS have another objective using it, which is to explain a certain term. In this case, the bundle appears in the middle of a sentence. The fact that the NNS insist on using the term instead of other alternatives such as *that is* and *or* seems to be related to their overdependence on some familiar formulas. This is also demonstrated in the use of *result* bundles. *As a result of* is a bundle commonly seen in academic prose. To avoid repeatedly using the same bundle, many writers often choose the alternatives. In adopting them, the NS and the NNS show a difference as revealed here. The NNS use *in the wake of* 3 times and *as a result the* 5 as alternatives of *as a result of*. By contrast, the NS choose *in the wake of, in the aftermath of* as its equivalents, 2 and 7 times each. Considering that *as a result the* is of mere a slight syntactic variation in sentence, it is clear that the NS are taking an advantage of a varied use of synonyms.

4.6 Others

Regarding the rest of the bundles that does not belong to anywhere, it can be said that a far varied types of bundles are used by the NS. The proportion of use by the NS is four times higher than the NNS. In addition to this, they seem freer in using bundles on comparison including *qualitatively as well as, as well as quantitatively*, and *as well as the.*

5. CONCLUSION

This comparative study has unfolded the differences between the NS and the NNS academic writing by comparing published economic journals. Through structural comparisons, salient differences were observed in the way they use lexical bundles.

First, as literatures suggest, the NS place more weight on NP-based bundles whereas the NNS have a tendency to rely on VP-based ones. In using VP-based bundles, the NS employ a higher number of and a wider variety of epistemic markers than the NNS, which has also been verified by the lack of *it is possible that* and *it is difficult to* in the NNS subcorpus. When introducing results and discussion, NNS are more reluctant to use the *we* + verb form now that learners traditionally are less likely to signal the ownership of the results and conclusions presented. In addition, the NS use a wider range of modifying bundles. Along with these, the NS tend to adopt the bundles involving certain words to a greater extent and with higher variation than the NNS. They also display a vast number of quantifiers. In addition, they are more likely to direct readers along the text by employing the related bundles. Lexical teddy bear is a phenomenon seen in the use of PP-based bundles in the NNS' texts. They have a tendency to cling to certain common terms than the alternatives. Finally, the NS seem freer in employing comparison bundles.

If I were to single out one future direction in which to continue this line of research, I would point to qualitative analyses of context. As demonstrated in the *in other words the* example, contextual analysis could reveal rather different usage of lexical bundles, because they do not use the same bundle in equally comparable situations. Furthermore, it could be revealing to examine lexical bundles that are more or less equally frequent in both groups now that the usage could prove the difference. Thus, a qualitative approach might be useful in exploring the usage of lexical bundles by the NS and the NNS.

References

Ädel, A., Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes, 31*(2), 81-92.

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word combinations. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications* (pp.101-122). Oxford: Oxford University Press.

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

Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes, 5*(2), 97-116.

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). *Longman grammar of spoken and written English.* London: Longman.

Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology, 14*(2), 30–49.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397-434.

Fletcher, W. (2007). KfNgram Version 2010. [Computer software]. (http://kwicfinder.com/kfNgram/ kfNgramHelp.html).

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: a study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics, 4*(2). 237-260.

Hoffman, C. (2000). The spread of English and the growth of multilingualism with English in Europe. In J.Cenoz & U.Jessner (Eds.), *English in Europe: The Acquisition of a Third Language (pp.1-21).* Clevedon: Multilingual Matters.

Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*(1), 41–62.

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

Lee, S. (2012). A study on the use of lexical bundles in science journals (과학저
널에 나타난 어휘다발의 사용 연구). Korean National Research Foundation.

Lindquist, H. (2009). *Corpus linguistics and the description of English.* Edin-
burgh University Press.

Romer, U. (2009). English in academia: Does nativeness matter? *Anglistik: In-
ternational Journal of English Studies, 20*(2), 89–100.

Salazar, D. (2011). Lexical bundles in scientific English: A corpus-based study
of native and non-native writing. (Unpublished doctoral dissertation).
Universitat de Barcelona, Spain.

Stubbs, M. (2007a). An example of frequent English phraseology: Distribution,
structures and functions. In R. Facchinetti (Ed.), *Corpus Linguistics 25
Years On* (pp. 89–105). Amsterdam: Rodopi.

APPENDIX A. Lexical bundles by the NS according to structural categories from LGSWE (Biber et al., 1999)

NP-based (258 types/40%)	PP-based (160 types/24%)	VP-based (209 types/32%)	Others (25 types/4%)
1) NP with "of" phrase fragment (137 types)	**1) PP with "of" phrase fragment (75 types)**	**1) Copula *be* + NP/ AdjP (56 types)**	but in fact the
the effect of the	of the effects of	is in the direction	but in retrospect it
the effects of the	as a result of	the dependent variable	half as large as
the size of the	of the effect of	is	as i noted earlier
the end of the	in the absence of	variable is the change	as well as the
the number of weeks	in the case of	all variables are in	half as large as
the timing of the	to the size of	also consistent with	a few days later
goals and methods of	in terms of the	the	early in the day
the contribution of the	in the direction of	are consistent with the	few days later on
the onset of the	in the form of	is consistent with the	higher than in the
the total quantity of	of the distribution of	is also consistent with	if the spread between
the impact of the	by the end of	consistent with the	larger than the effect
the level of the	in the aftermath of	lower	more than a few
the nature of the	on the order of	consistent with the	than half of the
the number of house-	in a number of	present	than that it amounted
holds	over the course of	consistent with the	than the effect esti-
the number of persons	beyond the scope of	view	mated
the number of series	for much of the	is the same as	a few years later
the overall effect of	in the course of	are the same as	as large as the
the panel study of	in the presence of	is the change in	as we discuss later
the presence of a	in the share of	is the ratio of	as well as for
the problem of identi-	in the wake of	too big to fail	as well as quantita-
fying	on the basis of	a data are for	tively
the response of the	after the end of	is highly statistically	of the figure shows
the results of the	as the sum of	significant	qualitatively as well as
the results of this	at the beginning of	a highly statistically	the results obtained
the role of expectations	at the end of	significant	when
the sensitivity of the	at the start of	and statistically sig-	when the survey year
the shape of the	at the time of	nificant at	
the stability of the	by the number of	statistically significant	
the start of the	from a variety of	and in	
the third panel of	from the end of	response is significant	
the time of the	in the context of	at	
the vast majority of	in the event of	are also very similar	
the wake of the	in the face of	are similar to those	
		similar to that of	

NP-based (258 types/40%)	PP-based (160 types/24%)	VP-based (209 types/32%)	Others (25 types/4%)
the weakness of the	in the number of	similar to those in	
values of the old	to the value of	is not surprising given	
the lower end of	in the spring of	is perhaps not surpris-ing	
a complete set of	of changes in the	available from the	
the bottom panel of	of the decline in	authors	
the fraction of the	of the first four	are available from the	
the present value of	of the initial effects	are close substitutes	
the value of the	of the number of	for	
a full set of	of the rise in	changes is relative to	
estimates of the effects	of this paper is	is a function of	
number of weeks of	on a base of	is a measure of	
that the effects of	as a function of	is from a separate	
that the size of	as a measure of	reported coefficient is	
the initial effects of	as the number of	from	
the magnitude of the	as the ratio of	of this paper is	
the size of this	at the expense of	too late for the	
the top quintile of	at the height of	the first is the	
a measure of the	by the ratio of	a both series are	
and the number of	for most of the	about half as large	
left panel of figure	in part because of	achieving at or above	
linear combinations of the	in the distribution of	are not subject to	
that much of the	of a break in	be the case that	
the first quarter of	of the degree of	been out of work	
the growth of the	of the difference be-tween	is a full set	
the growth rate of	of the group of	but not truly perma-nent	
the ratio of the	of the previous section	unable or unwilling to	
the reading achieve-ment of	of the relationship between	is beyond the scope	
the right-hand side of	of the size of	is in the top	
the scope of this	of time devoted to	is possible that the	
the top panel of	on a variety of	is the impact on	
the top quartile of	on other types of	is the sum of	
the value of their	on our measure of	not very sensitive to	
a function of the	on the part of	small relative to total	
a small number of	to some of the	standard errors are in	
an effect of about	to that of the		
an effect of those	to the end of	**2) VP with active verb (52 types)**	
and calculated effects of	to the question of	this would imply an	
cumulative effect of all	to the rest of	we now turn to	
cumulative effect of the	to the sum of	estimate the effects of	
	with the exception of		

NP-based (258 types/40%)	PP-based (160 types/24%)	VP-based (209 types/32%)	Others (25 types/4%)
different components of the	with the number of	measure the effects of	
one of the most		we do not find	
panel of the figure	**2) PP without "of" phrase fragment (85 types)**	we use a two-day	
panel of the table		see the online appendix	
scope of this paper	to the extent that	identify the effects of	
size of the effects	in the direction pre-dicted	overstate the effect of	
statistically significant effects of	in the bottom panel	data have generally found	
the beginning of the	in the online appendix	exceed some ten to	
the commitment of the	in this section we	indicate periods for	
the course of a	on the other hand	which	
the end of regular	at the same time	it amounted to several	
the experience of the	on the change in	measure the effect of	
the first half of	by the fact that	not seem to be	
the height of the	over the past five	now turn to the	
the likely effects of	as a decrease in	one might have ex-pected	
the rest of the	by the alternative the	see text for details	
the second quarter of	except where stated	see the text for	
the statistical signifi-cance of	otherwise	should have led to	
	in the present paper	the figure also shows	
the sum of the	under the alternative hypothesis	the table shows the	
the total number of	up to scale and	thus should have caused	
those of the present	for several reasons first	use data from the	
top panel of table	for this reason we	may have been a	
a better measure of	in online appendix table	as follows section i	
a great deal of	in the baseline specifi-cation	come as a surprise	
a large majority of		does not change the	
a larger fraction of	in the first two	fall behind on their	
a simple model of	in the top quintile	had a significant im-pact	
a small portion of	on the right-hand side	has a value of	
a small share of	over the indicated period	has changed over time	
a subset of the		has little impact on	
a share of all	to an increase in	have no effect on	
a wider range of	under the assumption that	one would expect to	
and the size of	with a complete set	we return to this	
bottom left panels of	with no prior account-ability	paper proceeds as fol-lows	
bottom right panel of	with respect to the	see online appendix table	
common components of the			
each of the four			
effect of the recent			

NP-based (258 types/40%)	PP-based (160 types/24%)	VP-based (209 types/32%)	Others (25 types/4%)
effects of changes in	with the view that	test for any x	
fifth column of table	at the longest and	tests the null hypoth-	
important part of the	due to declines in	esis	
number of persons in	during and after the	the bottom panel	
percentile of the	even in the absence	shows	
change	except where noted	the top panel shows	
the common compo-	otherwise	they would not be	
nent of	for a break in	those that did not	
the behavior of the	for the first time	use the baseline speci-	
that the value of	for the full sample	fication	
the average duration of	in a narrow window	we do not include	
stability of the factor	in many important	we do this to	
remainder of the paper	respects	we use three external	
right-hand side of	in response to the	what was then an	
equation	in the common compo-	would correspond to an	
an estimate of the	nent	would expect to see	
problem of identifying	in the same way	would need to be	
the	in the short run		
a number of different	off the balance sheet	**3) Anticipatory *it* +**	
the cross-sectional	on a weighted basis	**VP/adjP + (comple-**	
average of	other than that it	**ment-clause) (6 types)**	
the degree of leverage	over the coming year	it is important to	
the depth of the	rather than the longest	it is difficult to	
the effects of a	to the fact that	it is possible that	
the effects of changes	to those in the	it is not clear	
the emergence of the	with the present paper	it is reassuring that	
the estimated effects of	along the intensive	it seems reasonable to	
the event of a	margin		
the exception of the	and to what extent as	**4) Passive verb + PP**	
the fall of the	a first approximation	**fragment (50 types)**	
the first of these	as a proxy for	can be pinpointed to	
the fourth quarter of	as a result we	computed as described	
the fraction of time	as a whole in	in	
	as a whole the	as discussed in section	
	at about the same	defined as in figure	
2) NP without "of"	by corresponding in-	as can be seen	
phrase fragment (121	creases in	is determined at the	
types)	by the full faith	reported by the new	
the change in the	for a more detailed	results are reported in	
the correlation between	for any x less	that can be pinpointed	
the	for more than a	would have been pre-	
the difference between	for the changes in	dicted	
the			

NP-based (258 types/40%)	PP-based (160 types/24%)	VP-based (209 types/32%)	Others (25 types/4%)
a change in the	for those with few	are defined as in	
a decrease in the	for those with low	are reported in the	
an increase in the	for two reasons first	are summarized in table	
statistical significance at the	in addition to the	as described in footnote	
this up to the	in an effort to	as described in figure	
a point in time	in the literature for	as described in the	
the decline in the	in the past two	as reported by the can	
the extent to which	in the same direction	be seen in	
a cubic in the	in the second half	as measured by the	
a one-sided test in	in this case the	be accounted for by	
an effect on the	in two ways first	have been made in	
effects in the literature	on the same day	heavily involved in the	
little effect on the	over the past few	can be used to	
one-day event window	over time in the	were used to estimate	
from	over time using the	would have been ex-pected	
percent change in aver-age	per year or less	adjusted over time us-ing	
percent level for a	to a decrease in	adjustment is made for	
the factors at the	to what extent the	and are adjusted for	
the online appendix for	whether and to what	are based on the	
the results in the		are measured as the	
the rise in the		attributed to differ-ences in	
the text for details		be seen in table	
a large effect on		been driven by the	
a recent study by		both periods are pooled	
a shift in the		given the size of	
decline in hours per		is assumed to be	
propensity to consume out		is attributed to differ-ences	
that an increase in		is defined as in	
the degree to which		is thought to have	
the fact that the		is based on the	
the hypothesis that the		reported in the bottom	
the second is the		series are seasonally adjusted	
the second term on		variable is replaced with	
the sharp rise in		was driven by the	
a one-sided test the		weighted by the num-ber	
a single day or			
an independent influ-ence on			
changes in the relative			

NP-based (258 types/40%)	PP-based (160 types/24%)	VP-based (209 types/32%)	Others (25 types/4%)
changes over time in		would be allowed to	
correlation between		have been predicted to	
the two		expected to fall by	
cumulative percentage		given the importance	
change in		of	
its target for the		taken at face value	
other factors affecting			
the			
percent level and the		**5) (VP +) *that*-clause**	
percent per year in		**fragment (17 types)**	
percentage point per		these results show that	
year		our analysis is that	
percentage points per		our results show that	
year		could argue that the	
seasonally adjusted		the idea was that	
monthly data		the idea is that	
set of interviews from		the results indicate	
single day or two		that	
size and composition		we find that the	
using		we show that the	
some ten to twenty		reflect the fact that	
statistically significant		reports the results	
impact on		obtained	
term on the right-hand		that are subject to	
the changes in the		that did not adopt	
the decline in hours		that did not have	
the decline in poverty		that implies a real	
the effects in table		one could argue that	
the extent that the		that it amounted to	
the joint change in			
the past few years		**6) (V/Adj +) *to*-clause**	
the rate at which		**fragment (23 types)**	
the top left panel		multiplier is likely to	
what one might call		is likely to be	
a break in the		are likely to be	
a closer look at		were more likely to	
a permanent increase		more likely to be	
in		is more likely to	
a reduction in the		likely to have a	
a significant impact on		are less likely to	
a substantial increase		are most unlikely to	
in		somewhat likely to fall	
		somewhat or very likely	

NP-based (258 types/40%)	PP-based (160 types/24%)	VP-based (209 types/32%)	Others (25 types/4%)
an important role in an increase in a an indicator variable for and a cubic in and a half years all of the information and changes in the and the change in and the text for appendix for further details baseline specification described in basis points in the change in the slope degree to which the errors based on four large effect on the online appendix for further only a small fraction or a change in percentage point more than percentage points more than percentage points to the points approximately percent changes possible explanations for the predicted value from the present evidence that the rate at which the rates at the zero right and bottom left space spanned by the specification described in note		likely to be small appears to have been to estimate the factors to measure the effects to fall behind on allowed to engage in in order to maintain me to overstate the to become heavily involved to exceed some ten to identify the effects to overstate the effect **7) Others (5 types)** but it is not so we do not following the introduction of leading up to the using data from the	

NP-based (258 types/40%)	PP-based (160 types/24%)	VP-based (209 types/32%)	Others (25 types/4%)
specification that al-lows for			
the factors through the			
standard errors based on			
the baseline specifica-tion described			
the bottom right panel			
the conditions under which			
the past two decades			
the possibility that the			
the predicted value from			
the first principal component			
the first two columns			
the growth in the			
the remainder of the			
the results in table			
the results in this			
the space spanned by			
the top right and			
those in the bottom			

APPENDIX B. Lexical bundles by the NNS according to structural categories from LGSWE (Biber et al., 1999)

NP-based (68 types/ 25.7%)	PP-based (63 types/ 23.8%)	VP-based (127 types/ 49%)	Others (4 types/ 1.5%)
1) NP with "of" phrase fragment (45)	**1) PP with "of" phrase fragment (32)**	**1) Copula *be* + NP/ AdjP (23)**	as well as the
the null hypothesis of	as a percentage of	economy is more likely	as long as the
the size of the	in the form of	is a group of	which all four vari-
the fundamentals of the	of rule of law	numbers in parenthe-	ables
the rest of the	of this paper is	ses are	a positive and signifi-
a small number of	as one of the	all four variables are	cant
and the level of	in the context of	are consistent with the	
estimated coefficients of the	on the basis of	are more sensitive to	
a deeper level of	in terms of the	is a function of	
and other measures of	at the end of	is equal to the	
and the rule of	for the period of	is higher than that	
important driving forces of	on the level of	is obvious that the	
the core elements of	about the value of	is the same as	
the end of the	in the case of	is one of the	
the existence of a	on the value of	is the measure of	
the passage of the	as a proportion of	the dependent variable	
the positive impacts of	as a result of	is	
the steady-state level of	as a sub-pillar of	and the other is	
a full range of	during the period of	of the paper is	
a high level of	for the purpose of	an economy can be	
a new way of	in the field of	can be an important	
a part of the	in the group of	can derive the following	
a plethora of studies	in the period of	can serve as a	
an increasing number of	in the wake of	more open to the	
and the exposures of	of no serial correlation	accounting for more than	
findings of previous studies	of the null hypothesis	and significant at the	
the negative sign of	on the concept of		
that the effects of	on the rule of	**2) VP with active verb (22)**	
that the number of	on the use of	not seem to be	
that the rule of	on rule of law	note the numbers in	
that the size of	as rule of law	denote statistical sig-	
	to the change of	nificances at	
	with the share of	we can derive the	
		do not seem to	
	2) PP without "of" phrase fragment (31)		

NP-based (68 types/ 25.7%)	PP-based (63 types/ 23.8%)	VP-based (127 types/ 49%)	Others (4 types/ 1.5%)
that the value of	on the other hand	does not seem to	
the animal spirits of	in appendix i except	we group firms by	
the characteristics of	in the sense that	we have the following	
the	at the same time	we look at the	
the common factor of	in other words the	derive the following	
the determination of	per capita in thousands	expression	
the	in this paper we	do not appear in	
the direction of causal-	as a proxy for	examine the relation-	
ity	as a result the	ship between	
the first set of	in the short run	have a negative rela-	
the number of observa-	among the selected	tionship	
tions	variables	i have used the	
the number of parties	as an independent	increased the volatility	
the relative price of	variable	of	
the relative size of	as opposed to sustain-	not appear in the	
the results of the	able	play an important role	
the second moments of	in the long run	show that of the	
the value of nature	on the relationship	table # summarizes the	
the share of the	between	the framework act on	
	with respect to the	vary depending on the	
2) NP without "of"	as a tool to	we focus on the	
phrase fragment (23)	as the following		
that an increase in	for convenience we will	**3) Anticipatory *it* +**	
the numbers in paren-	for this purpose we	**VP/adjP + (comple-**	
theses	in the near future	**ment-clause) (8)**	
an important role in	in the next section	it could be hypoth-	
the relationship be-	in the previous section	esized	
tween the	in this regard the	it is expected that	
all countries in appendix	in this section we	that there is a	
the solution to the	into conformity with	it is clear that	
and percent changes in	the	it is not clear	
changes in the funda-	on the same or	it seems likely that	
mentals	over the sample period	it is important to	
general and bilateral	to this end a	it is obvious that	
exemptions	to this end the		
rule of law the	with the previous stud-	**4) Passive verb + PP**	
the sense that it	ies	**fragment (43)**	
a channel through		referred to as the	
which		as seen in the	
a positive effect on		paper is organized as	
a positive long-run		is determined by the	
relationship		given by the following	
an increase in the		organized as follows	
model in this paper		section	

NP-based (68 types/ 25.7%)	PP-based (63 types/ 23.8%)	VP-based (127 types/ 49%)	Others (4 types/ 1.5%)
null hypothesis at the		are expected to be	
only a small number		can be interpreted as	
relationship among the		are explained by the	
variables		are measured in sepa-	
rest of the paper		rate	
stability and distance		are obtained from the	
from		are offset by the	
the other control vari-		as shown in the	
ables		can be defined as	
the short and long-run		can be seen as	
		cannot be rejected for	
		four variables are	
		cointegrated	
		is defined as the	
		is given by the	
		the paper is organized	
		used as a proxy	
		was hit by the	
		are assumed to be	
		are collected from the	
		are effectively limited	
		by	
		are found to have	
		are given in parenthe-	
		ses	
		are not related with	
		are positively exposed	
		to	
		are reported in table	
		are shown to be	
		as shown in table	
		as stated in the	
		be considered as a	
		can be offset by	
		considered in this	
		study	
		errors are given in	
		is thought to be	
		is used as a	
		is documented in ap-	
		pendix	
		more integrated with	
		the	

NP-based (68 types/ 25.7%)	PP-based (63 types/ 23.8%)	VP-based (127 types/ 49%)	Others (4 types/ 1.5%)
		results are reported in this paper is organized **5) (VP +) *that*-clause fragment (11)** sense that it is i have argued that it shows that the the results show that the reason is that observers have argued that we assume that the we can see that this implies that the we can argue that assume that the market **6) (V/Adj +) *to*-clause fragment (19)** this paper is to this study is to is more likely to likely to invest in are more likely to are likely to be likely to form pessimistic market likely to be more likely to form more likely to invest to accede to the to be an important to be the most is to examine the to have a positive important to note that seem to be a seems to be the in order to gain **7) Others (1)** if we substitute equation	

ABSTRACT

Lexical Bundles in Economic Research Articles by Native and Non-native Speakers of English

Bit-Na Choi

This paper examines the use of lexical bundles in economic research articles of native and non-native speakers of English. Similar and different structural features are compared between the articles of American and Korean authors, from the Brookings Institution and the Korea Institute for International Economic Policy (KIEP), which amount to 262,653 words. KfNgram (Fletcher, 2010) was used to retrieve corpus query of four-word lexical bundles. Structural analysis was conducted following the method by Biber et al. (1999). Suggestions for future research and improvements are provided.

Key Words   lexical bundles (LB), research articles, economic research articles, academic writing, native speakers of English, non-native speakers of English, structural analysis