Abstract

The present study is a corpus-driven investigation on frequency, structure and function of lexical bundles (LBs) in linguistic research articles of native speakers of English (NS) and non-native speakers of English (NNS). By the use of *Antconc 3.4.3* (Anthony, 2014), 4-word LBs were retrieved from 909,259 words of corpora that consist of 91 research articles published within the last 11 years from leading journals in theoretical linguistics, 52 NS researchers and 39 NNS researchers. In retrieval, the LBs were filtered by frequency and dispersion criteria. This initial list was filtered again by applying a set of exclusion criteria. In this manner, the most frequent 20 LBs used by each group were examined. The structural and the functional features of the generated LBs were explored through concordance analysis by modified versions of Biber et al.'s (1999) structural framework and Hyland's (2008a) functional taxonomy.

This mixed method approach of quantitative and qualitative analyses revealed NNS' more frequent use of LBs than the NS counterpart, which originated from more common use of metadiscursive bundles and less variations in use. In structure, both groups showed the similar overall tendency to favor PP-based, VP-based and NP-based LBs in order. Notably, the NNS preferred passive forms than active, which points to NNS' reluctance to reveal the authorial identity (Hyland, 2008b). Functionally, although the NS and the NNS demonstrated a similar pattern to favor LBs in the order of text-oriented, research-oriented and participant-oriented, there were several noteworthy findings. In conveying personal views, the NS used indirect forms including anticipatory-*it* structures, whereas the NNS adopted personal pronoun *I/we*. Such self-mention strategies suggest that predicative adjectives might not be a familiar option for the NNS. Contrary to the literature, the NNS exhibited a varied use of hedges. The diversity in the use of the "native bundles" (Hyland, 2000b) ranging from the

high end of certainty to the lower ends (Biber, 2006) indicates their expertise, while the extensive uses of modal and “speech act” verbs indicate their nonnative background.

The present thesis aligns with the previous studies in revealing the differences in use of LBs between NS and NNS. However, the newly found similarities between the NS and the NNS researchers suggest that expertise is also an important element as much as nativeness.

Keywords : corpus, lexical bundles, native speakers of English (NS), non-native speakers of English (NNS), linguistics, research articles

Student Number : 2011-23083

Table of Contents

List of Tables

1. Introduction

1.1 Background and Motivation

Exactly how different are native speakers from non-native speakers? Given the education, development and worldwide spread of the English language, what are the literary traits that distinguish native speakers from non-native speakers? Throughout changes in history, many parts of the world underwent British and North American influence, the World Wars and globalization. English is used daily today in communication and academia by many non-native speakers of English (hereafter "NNS") around the world (Hoffman, 2000). Given its status as *lingua franca* (Hoffman, 2000), English is relevant to all fields in life and business, but perhaps it is most prominent in academia, where conventional high-proficiency English is required. Countless research is conducted by NNS, as dramatically higher number of academic texts is being produced by this group (Römer, 2009). As non-native speakers become more fluent, a multitude of studies have questioned the elements of what makes a well-written prose (Salazar, 2011). At the center of the discussions are frequent word combinations, which have been regarded as an indicator of proficient language, including academic writing (Cortes, 2004).

Frequent word combinations have been employed in various forms by scholars. They are also referred to as *clusters* (Hyland, 2008a), *recurrent word combinations* (Altenberg, 1998; Lindquist, 2009), *multi-word expressions* (Sag et al., 2002), *n-grams* (Stubbs, 2007a), and *lexical bundles* (Biber & Barbieri, 2007). Throughout this study, *lexical bundle* is the primary term referring to recurrent word sequences that occur in the academic register, as has been used by Biber in his series of studies upon which the theoretical framework of the present study is based.

In the last three decades, scholars' use of lexical bundles received not less attention. Discipline-specific and cross-disciplinary studies have been conducted by various researchers. Hyland (2008a), one of the pioneering researchers, observed that expertise in the language mattered more than being a native speaker of it. Salazar (2011) also carried out a study to

explore the differences found in the use of lexical bundles by scientists whose first language (hereafter L1) was Spanish and scientists whose L1 was English. She concluded that certain variations existed such as the overuse of bundles by NNS and the restricted use of participant-oriented bundles by NS. Surprisingly, no other discrepancies were found.

In Korea, Chung and Song (2012) examined different uses of lexical bundles by L1 Korean-speaking and L1 English-speaking scholars in the Department of English Literature, English Linguistics, and English Education. Korean scholars adopted more varied types of lexical bundles. Functionally, referential bundles occupied the majority of the NS lists, and text discourse organizers were most frequently found in the NNS lists. Different selections in stance expressions were also observed. Although the study was an extensive effort that involved cross-disciplinary research of Korean scholarly works, a qualitative analysis could have offered a more comprehensive picture.

Therefore, the present study attempts to expand the small body of existing literature by further exploring this issue with a mixed method approach of quantitative and qualitative analyses. The current study examines how the use of English lexical bundles differs between L1 Korean-speaking researchers and L1 English-speaking researchers. This is done by studying the most common 4-word lexical bundles from linguistic journals employed by NS and NNS researchers. The thesis also investigates the structural characteristics of the identified bundles in the NS and the NNS lists. The study concludes by discussing the functional characteristics of the bundles used by the NS and the NNS.

1.2 Research Questions

The differences in lexical bundles in linguistic articles used by native and non-native English speakers are addressed in this study. This study intends to answer the following research questions:

1. What are the most common 4-word lexical bundles in linguistic research articles adopted by native and non-native English speakers? What differences do they show?
2. What are the structural characteristics of lexical bundles in linguistic research articles? Are there any structural differences between the lexical bundles used by native and non-native English speakers?
3. What are the functional characteristics of the lexical bundles in linguistic research articles? Are there any functional differences between the lexical bundles used by native and non-native English speakers?

2. Review of Literature

This chapter presents a review of previous studies regarding lexical bundles use. It begins with a description of lexical bundles, where the definition, characteristics, retrieval criteria, and functions are mentioned. The next section discusses some of the studies on lexical bundles use in academic literature. We look at Biber et al.(1999)'s study that triggered the following studies of lexical bundles in the genre of academic prose. We next review the studies on the bundles used by L1 English-speaking and second language (hereafter, L2) English-speaking researchers.

2.1 Lexical Bundles: What are They?

The notion of lexical bundles was first introduced by Biber et al. (1999) in their comprehensive study of English grammar, the *Longman Grammar of Spoken and Written English* (LGSWE). In the book, lexical bundles were defined as "bundles of words that show a statistical tendency to co-occur" (p. 989) and as "recurrent expressions, regardless of their idiomaticity, and regardless of their structural status" (p. 990).

Biber and Barbieri (2007) identified three major characteristics of lexical bundles that distinguish them from other types of word combinations. First, lexical bundles are extremely common, as the definition suggests. Second, most of them are not idiomatic in meaning and not perceptually salient. They are transparent in meaning and fully derivable from individual words which comprise the whole bundle, as seen in *one would indicate that*, *can be said that*, and *it can be seen*. Finally, they do not necessarily represent a complete structural unit. Rather, most of them function to bridge two structural units. They begin at a boundary of a clause or a phrase, and the final words of the bundle are the beginning components of a second structural unit. Most bundles in speech bridge two clauses (e.g. *well that's what she, you want to know*), while those in writing bridge two phrases (e.g. *in the presence of, in the case of*). It was also demonstrated in Biber et al. (1999) that most bundles in conversation are clausal whereas those used in academic prose are phrasal.

These lexical bundles have two important criteria in its identification: frequency and dispersion (Ä del & Erman, 2012). These two have been provided in literature as the key to generate lexical bundles by using corpus tools. The cut-off frequency is the number of lexical bundles which need to be included in the analysis. The frequency cut-off point has somewhat different thresholds for spoken and written corpora. For spoken data, usually a higher bar has been applied, since spoken corpora tend to produce more lexical bundles as shown in Biber et al. (1999). Biber and Barbieri (2007) employed 40 times per million words (hereafter PMWs) for spoken data. As for written corpus, however, a lower cut-off point has been suggested, usually varying from 10 to 40 PMWs. Studies have adopted 10 PMWs (Biber et al., 1999; Simpson-Vlach and Ellis, 2010), 20 PMWs (Cortes, 2004; Hyland, 2008b), 25 PMWs (Chen & Baker, 2010; Ä del & Erman, 2012), and 40 PMWs (Biber et al., 2004). Dispersion criterion determines the range in which the lexical bundles should occur across different texts. This is to avoid idiosyncratic uses of individual authors. The range distribution has often been set at 3 to 5 texts, sometimes 6. For example, Biber et al. (1999) opted for 6 texts while Simpon-Vlach and Ellis (2010) established it at 3 out of 4 academic divisions. Chen and Baker (2010) required it to be 3, which was followed by Ä del and Erman in 2012. Hyland (2008b) adopted a rather conservative approach, ranging across 10% of texts. In the study of lexical bundles, length has been one of the primary concerns. 3 to 6 word units have been mainly studied thus far, but the length that has mostly been researched is 4-word units, especially in studies using written corpora. This can be explained by the fact that 3-word units are more commonly produced than 4-word units and that a number of them are usually incorporated in the 4-word bundles. 5-word and 6-word units are somewhat rarer to examine (Biber et al., 1999).

Although lexical bundles do not represent complete structural units, they are still considered as “important building blocks in discourse” (Biber & Barbieri, 2007, p.270). Scholars attempted to designate the functions, beginning with Biber et al. (2004). They classified primary functions of lexical bundles into the following three categories: stance expressions, discourse organizers, and referential expressions. Stance bundles indicate attitudes or assessments of certainty that frame other propositions. Discourse organizers link the preceding and the following propositions. Referential bundles directly refer to physical or

abstract entities, or to the textual context itself, either to identify the entity or to pinpoint a particular attribute of the entity as something important (Biber et al., 2004). However, Hyland (2008b) modified the classification of bundles for academic written corpora. He classified bundles into three broad functions of research-oriented, text-oriented, and participant-oriented and then divided each into subcategories.

2.2 Lexical Bundles in Academic Writing

For the reasons aforementioned, academic prose has captivated the attention of numerous researchers over the past decades. Among all, Biber et al. (1999) are noted for the pioneering study on lexical bundles in the genre of academic writing. Lexical bundles that represent complete structural units are shown to be nearly non-existent in academic prose. If a bundle is structurally complete, it is usually a prepositional phrase that functions as discourse markers. Over 60% of the recurrent expressions in academic prose were parts of noun phrases or prepositional phrases. Furthermore, most of the bundles end in a function word, as in *as a result of* and *in the case of*.

The topics of having expertise versus being a native speaker have been prevalent for decades. Hyland (2008a) is one of the great researchers and a significant contributor to the literature. He broadened the scope of discipline-specific research for cross-disciplinary study. The disciplines selected were electrical engineering, business studies, applied linguistics, and microbiology. As for data, research papers of native speakers were targeted. There were also L1 Cantonese speakers in two groups: PhD students with PhD dissertations and students in master's program with master's theses. Both student groups favored anticipatory-*it* structures in which they could disguise writer identity. They also adopted more bundles that served to engage readers. However, PhD students had some similarities with the NS writers. More text-oriented bundles appeared in their texts contrary to the master's students who included a high number of research-oriented bundles. The NS writers employed a multitude of text-oriented bundles, with only a few research-oriented bundles.

Salazar (2011) followed the thread of discipline-specific research by investigating the uses of lexical bundles in scientific research articles by NS and NNS writers. A collection of research articles in biology and biochemistry by native English speakers were compared to biomedical research articles by native Spanish-speaking scientists. As had been demonstrated in the literature, Salazar observed that NNS overused certain bundles which resulted in unnecessary repetitiveness and a lack of variation. Use of participant-oriented bundles was particularly restricted, which may indicate NNS's limited awareness of their usage and importance. Furthermore, there was a somewhat surprising finding that there was no significant difference between the two groups of writers, contrary to previous research. She concluded that the control for the goal, the degree of expertise, topic, text type, and author profile could be important factors for comparative study.

Over the years, many studies on lexical bundles took place in Korea. All of them were carried out in a setting of English as a Foreign Language (hereafter, EFL), mostly with academic prose as the focus of study. Lee (2012) compared the differences in use of lexical bundles of NNS Korean scientists to those of NS scientists in research articles of life science. He found that NNS overused certain bundles while neglecting to use hedging devices. Certain bundles from the NNS corpus exceeded the NS in frequency or type, including past tense passive forms and the personal pronoun *we*. He explained the tendencies by the NNS' devotion to empirical procedure.

In this context, Chung and Song (2012) examined the use of lexical bundles by NNS Korean scholars in the fields of English literature, English linguistics, and English education and compared them to those of the NS scholars. They found that NNS used more diverse range of lexical bundles, and fewer bundles were used in common by the NS and the NNS. The most bundles used by the NS were in English linguistics, while the NNS used most bundles in English education. Generally, in contrast to English literature which had the most bundles expressing time reference, English linguistics had the most bundles that presented the basis or comparison of arguments. In English education, the bundles reflected the empirical nature of the field and exhibited the most difference between the NS and the NNS. Although it was a pioneering attempt to study the lexical bundles used by L1 Korean English users in comparison

of L1 English users, quantitative analysis alone was inadequate to explain the observed patterns. To achieve the purpose, concordances checks could have been useful.

These previous studies have provided valuable insights into how NNS use lexical bundles in academia. In this strand of research, academic writing of Korean researchers in linguistics remains to be further explored and is discussed in the present study.

3. Data and Methodology

This chapter deals with data and the methodology used to analyze the data. It explains a description of data, the data collection process, and data analysis method. Exclusion criteria for lexical bundles are also provided, followed by structural and functional taxonomies for classification.

3.1 Data

Data for this study is a compiled collection of linguistic research articles taken from widely known leading journals, amounting to 909,259 words. Linguistics was selected as an object of study because little research had been conducted in the literature. The scope was narrowed to theoretical linguistics since it is the core subfield of linguistics that had long sustained it. Articles were collected from journals of theoretical linguistics whose selected subdisciplines included phonology/phonetics, syntax, and semantics/pragmatics. The subdiscipline of phonology was extended to include phonetics since certain overlaps existed and the boundary was blurry for a number of articles selected, which was also applied to the subfield of semantics and pragmatics.

The corpus is composed of 91 research articles published from the years 2004 to 2014. Table 1 gives an overview of the data used for each of the NS and the NNS subcorpora. The journal names from which the articles were selected are provided in Appendix C.

Table 1. The native-speaker and the non-native speaker subcorpora

	NS		NNS	
Subdiscipline	Texts	Words	Texts	Words
Phonology/Phonetics	19	218,446	17	137,199
Syntax	17	167,407	9	77,386
Semantics/Pragmatics	16	211,048	13	97,773
Total	52	596,901	39	312,358

Efforts were made to control the number and the size of texts of the NS and the NNS corpora, but due to the varied length and availability of texts, some variations in corpus size exist. Articles of theoretical linguistics from the NS corpus tend to be longer than those in the corpus of the NNS.

Close attention was made to identify the NS/NNS status of writers, and NS status was determined by referring to Swales (1985). Among the suggested criteria, last name and institutional affiliation were foremost considered. Of the NNS corpus, articles which were written by native speakers of English as coauthors were excluded.

Journals were selected from Seoul National University (hereafter "SNU") online library system. SNU subscribes to more than 172,000 online journals. In order for a journal to be purchased, the library committee undergoes a strict review of those recommended by professors and students. The journal purchase is determined based on its impact factor, the number of institutions that purchased it, and the existence of the hardcopy. Having been collected under such conditions, the online journal source of the SNU library was deemed reliable. The selected journal articles were downloaded from the online versions of journals and then converted into plain text files. The converted electronic files were saved in the names of the subfield, the author, the year of publication, the title and the journal name. To ensure accurate data processing and analysis, the files were cleaned of headers, footers, annotations, tables, diagrams, and references. After the files were cleaned of irrelevant data, the number of words in each file was counted and added in the name of the files.

3.2 Methodology

3.2.1 Lexical Bundles Identification

As for the target bundle, 4-word lexical bundles were considered in the study. The 4-word scope has been "the most researched length for writing studies" (Chen & Baker, 2010, p.32). Biber et al.(1999) has proven that 3-word units possibly lack significance due to its commonality while 5-word and 6-word units are somewhat rarer to conduct research. However, in a recent study of

Simpson-Vlach and Ellis (2010), 3-word lexical bundles were found to be the majority of important multi-word combinations. Nonetheless, since many 3-word lexical bundles are subsumed in the 4-word bundles, as in *in the context* in *in the context of,* 4-word became the focus in the current study.

In the literature, several criteria have been identified in retrieving lists of lexical bundles. The first criterion is the cut-off frequency, which sets the number of lexical bundles to be included in the analysis. Studies have ranged from 10 to 40 times PMWs (Biber et al., 1999; Hyland, 2008b; Chen & Baker, 2010; Ädel & Erman, 2012), and a minimum frequency of 20 times PMWs was established for this study following Cortes (2004). Dispersion criterion was also considered, to mark the range in which lexical bundles occur across different texts. They usually ranged from 3 to 5 texts in previous studies (Biber & Barbieri, 2007; Cortes, 2004; Chen & Baker, 2010), but a conservative approach took a minimum of 10% of texts (Hyland, 2008a). In this study, 3 texts were chosen. This method of setting the frequency at 20 times and dispersion at 3 texts provided an optimum size of the lexical bundles retrieved.

As a result, 4-word lexical bundles were generated that met the criteria, which were to occur at least 20 times PMWs and across more than 3 texts. As for a corpus tool, *Antconc 3.4.3* (Anthony, 2014) was used. In total, 377 bundles were produced.

3.2.2 Exclusion Criteria

After automatic retrieval of lists of 4-word lexical bundles, further work was carried out to ensure that the produced lists of lexical bundles represented relevant data. By referring to Salazar (2011), specific exclusion criteria were established. The following table displays a summary of the criteria.

Table 2. Exclusion criteria (adapted from Salazar (2011))

Categories	Examples
Topic-specific bundles	*the extension of a, jaw and the lips*
Bundles with number	*shown in figure 1*

Proper nouns	*the university of California, the two-way ANOVA*
Bundles composed exclusively of function words	*as in b the, as in a or, et al and the, of # with #*
Bundles of function words and topic-specific words	*of the speaker s; from the speaker s*
Examples	*don t have to, I think it s, don t know i*
Interviews	*I think it s, don t want to*
Meaningless bundles	*nan emm nan emm, et al s study*
Fragments of other bundles	*the scope of this, been shown to be, degree to which they*

Salazar's (2011) criteria were slightly modified for the current study. The newly added categories were *proper nouns, bundles composed of function words and topic-specific words*, *examples*, and *interviews*. *Proper nouns* has often been deleted in the previous studies (Chen & Baker, 2010; Ädel & Erman, 2012) since it carried less significance in research. Along with *topic-specific bundles*, *function words with topic-specific bundles* was removed, for it held no particular value without topic-specific words. *Examples* and *interviews* were removed since they did not appear in the corpora for the current study. Some categories in Salazar (2011) were combined with the existing categories in new criteria. *Bundles with random numbers*, *time bundles*, *temperature*, *volume* and *length* bundles had number in each and were combined as *bundles with number*. *Web noise* was not included since it had been deleted while the electronic files had been saved.

Other categories followed those of Salazar (2011). Topic-specific bundles that had a limited use in specific domains or subdisciplines were excluded. Examples such as *the subject of a, the extension of a, of the adaptation of* made it to the list for they had certain terminological meaning in contexts. Their primary meaning generally used was not in focus in the texts. *Bundles made up exclusively of function words* was deleted now that it held no meaning in the texts. Among the generated lexical bundles, *meaningless bundles*, such as *nan emm nan emm, et al s study*, was also removed from the lists. As for *fragments of other bundles*, semantic meaning was foremost considered. Bundles that carried more significant meaning and thus contributed more to the text were preserved. For example, when *the use of the* appeared 33

times and *in the use of* had 7 appearances, the latter was preserved while the former was reduced to 26.

3.2.3 Structural Classification

Following the exclusion process, the next step was to examine structural characteristics of the lexical bundles. For an objective and reliable analysis, a well-established classification of Biber et al. (1999) was adopted. Some modifications were made to this structural classification by referring to Salazar (2011), for it was more suitable for academic writings. Four new categories were added: VP with active verbs, verb phrases with personal pronoun *I/we*, other VP fragments, and adjectival phrases. They were grouped into six categories: NP-based, PP-based, VP-based, adjectival phrase, adverbial clause, and others. The modified taxonomy is presented in the following table.

Table 3. Modified structural classification (adapted from Biber et al., 1999)

Category	Pattern	Example
NP-based (Noun phrase with post-modifier fragment)	1) NP with "of" phrase fragment	*the use of the* *results of the present*
	2) NP without "of" phrase fragment	*the relationship between the* *difference between the two*
PP-based (Preposition + noun phrase fragment)	1) PP with "of" phrase fragment	*on the basis of* *in terms of the*
	2) PP without "of" phrase fragment	*in the previous section* *in this paper i*
VP-based	1) Copula *be* + NP/adjP	*is responsible for the*
	2) VP with active verb	*does not have a*
	3) Anticipatory *it* + VP/adjP + complement-clause)	*it should be noted* *it will be shown*
	4) Passive verb + PP fragment	*is followed by a* *is used with the*
	5) (VP+) *that*-clause fragment	*that there is no* *that the use of*
	6) (V/Adj+) *to*-clause	*was found to be*

	fragment	
	7) VP with personal pronoun *I/we*	*i have argued that* *we expect that the*
	8) Others	*which is taken from*
Adjectival phrase		*due to the fact*
Adverbial clause		*as shown in the*
Others		*as well as the*

In the modified classification, some of the categories in Salazar (2011) were not included. Unlike the original three subcategories of NP, *other noun phrase* was removed, for it could be assigned to either of the two remaining subcategories. *Other passive fragment* was excluded for the same reason. Instead, *active verb* was added to the VP subcategory.

3.2.4 Functional Classification

Another important framework in the analysis was functional taxonomy. The classification was originally adopted from Hyland (2008a), modified in accordance with Salazar (2011), and underwent slight modification. The table below illustrates a modified taxonomy by the researcher in referring to Salazar (2011). In accordance with the study, *topic bundles* was deleted. The *grouping* subcategory was newly created to include the bundles of categorization, classification, and grouping. *Citation* was added to classify those that cite sources. Hyland's (2008a) *contrastive* and *resultative* functions were subdivided by the narrower subcategories *additive* and *comparative*, and *inferential* and *causative*. However, for the present study, text-oriented generalization and participant-oriented acknowledgement bundles were not included due to their absence.

Table 4. Modified functional taxonomy (adapted from Hyland, 2008a)

Research-oriented bundles	**Text-oriented bundles**	**Participant-oriented bundles**
Help writers to structure their activities and experiences of the real world	Concerned with the organization of the text and its meaning as a message or argument	Focused on the writer or reader of the text

Location
Indicate place, extremity and direction
at the end of
Procedure
Indicate events, actions and methods
the use of the
Quantification
Indicate measures, quantities, proportions and changes thereof
the total number of
Description
Indicate quality, degree and existence
the distribution of the
Grouping
Indicate groups, categories, parts and order
as one of the

Additive
Establish additive links between elements
in addition to the
Comparative
Compare and contrast different elements
in contrast to the
Inferential
Signal inferences and conclusions drawn from data
can be interpreted as
Causative
Mark cause and effect relations between elements
the results of the
Structuring
Text-reflexive markers that organize stretches of discourse or direct the reader elsewhere in text
in the present study
Framing
Situate arguments by specifying limiting conditions
in the sense of
Citation
Cite sources and supporting data
is used with the
Objective
Introduce the writer's aims
in order to examine

Stance
Convey the writer's attitudes and evaluations
it is possible that
Engagement
Address readers directly
as shown in figure

Biber et al. (2004) observed that a bundle can serve varied functions in different contexts. They mentioned that bundles such as *the beginning of the* and *at the end of* can function not only as a time reference but also as place and text deictic references depending on the context. To determine the predominant function of a bundle in the presence of multifunctionality requires concordance examinations and the sorting of bundles according to the most common use. In this research, secondary functions were also considered. Most secondary functions were carried by participant-oriented bundles. For example, *we expect that the* had a stance function as a participant-oriented bundle but also served for an inferential function as a text-oriented bundle. Hence, the idea of multifunctionality produced 65 bundles that were multifunctional. Following

the steps described in this section, the identified bundles were assigned to aforementioned structural and functional taxonomies. The results were produced after concordance analysis, which will follow in the next section.

4. Data Analysis and Results

In this chapter, frequency and usage patterns of the identified lexical bundles in the NS and NNS corpora are analyzed within the fields of theoretical linguistics. The frequency patterns of lexical bundles are reviewed, followed by structural and functional characteristics. Due to different corpora sizes of the compiled collection, relative frequencies per 100,000 words were calculated to examine the overuse or underuse by one group over another.

4.1 Frequency of Lexical Bundles

4.1.1 Frequency Patterns of Lexical Bundles

Frequency patterns of lexical bundles in the NNS corpus were compared to the results from the NS corpus. The retrieved bundles from the NS corpus amounted to 107 while those from the NNS corpus were 270, each with 2,212 and 2,897 frequencies. Considering the inconsistent corpus sizes, the relative frequencies of the NS were 370.58 and those of the NNS were 927.46. The most frequently occurring 20 bundles in the NS and the NNS corpora were compared by raw frequency. When more than one bundle had the same raw frequency, the bundle with a wider range was ranked first. The results are presented in Table 5. The full lists of bundles can be seen in Appendices A and B.

Table 5. Top 20 lexical bundles in the NS and the NS corpora

Rank	NS	Raw freq.	Rel. freq.	NNS	Raw freq.	Rel. freq.
1	**On the other hand**	69	11.56	**On the other hand**	147	47.06
2	**In the case of**	60	10.05	As a function of	48	15.37
3	**On the basis of**	57	9.55	**With respect to the**	45	14.41
4	**With respect to the**	50	8.38	In the present study	45	14.41
5	To account for the	46	7.71	**In terms of the**	41	13.13
6	**The fact that the**	39	6.53	In this section i	34	10.88
7	The rest of the	39	6.53	**In the case of**	32	10.24
8	The left edge of	37	6.20	The results of the	28	8.96
9	In addition to the	35	5.86	In the next section	27	8.64
10	**As well as the**	33	5.53	The duration of the	27	8.64
11	**In terms of the**	33	5.53	**At the same time**	26	8.32
12	At the level of	33	5.53	The use of the	26	8.32

13	It is clear that	33	5.53	**The fact that the**	25	8.00
14	In this case the	32	5.36	**As well as the**	23	7.36
15	It is possible to	32	5.36	In other words the	23	7.36
16	That there is a	32	5.36	In the previous section	21	6.72
17	In a way that	30	5.03	We have seen that	21	6.72
18	**At the same time**	30	5.03	**On the basis of**	20	6.40
19	In the sense that	29	4.86	Prediction is borne out	20	6.40
20	At the end of	28	4.70	We expect that the	20	6.40

In Table 5, bundles shown in bold were shared by both groups from the top 20 lists. The commonly shared ones in the full lists are underlined. Their absence in the other corpus is a strong indicator of its overuse in one corpus. The examples of the overused and the underused bundles by the NNS are summarized in Table 6.

Table 6. Examples of lexical bundles overuse and underuse in the NNS corpus

Overused	on the other hand, as a function of, with respect to the, in the present study, in terms of the, in this section I, in the case of, the results of the, in the next section, the duration of the, at the same time, the use of the, the fact that the, as well as the, in other words the, in the previous section, we have seen that, prediction is borne out, we expect that the
Underused	on the basis of, to account for the, the rest of the, the left edge of, in addition to the, at the level of, it is clear that, in this case the, it is possible to, that there is a, in a way that, in the sense that, at the end of

A look at the overused bundles by the NNS revealed their tendency to employ a high number of metadiscursive bundles, as evidenced by *on the other hand, in the present study, in this section i, in the next section, in other words the, in the previous section, we have seen that,* and *we expect that the*. Such relatively frequent use of metadiscursive bundles by the NNS appears to be their distinctive trait. Other notable finding was in the underused pattern demonstrated by the NNS, which further detail a proposition and hedges, as in *it is clear that, in the sense that,* and *in a way that. It is clear that* was employed to show the authorial stance of a proposition to a certain degree or extent, while the other two served to qualify a following clause.

4.1.2 Summary

In this section, frequency and patterns of lexical bundles adopted by NS and NNS scholars were discovered. The revealed repetition of lexical bundles by NNS has been recognized as a problem of second language writers in writing academic prose (Pawley & Syder, 1983). Relative frequencies of lexical bundles disclosed the NNS writers' tendency to show "a greater awareness of the text as a text" (Ädel & Erman, 2012, p.86) by employing metadiscursive bundles. Previous studies found that NNS tend to show limited and restricted use of hedges (Ädel & Erman, 2012; Chen & Baker, 2010; Cortes, 2004; Hyland, 2008b; Salazar, 2011), which are also corroborated by the findings of the present study.

4.2 Structural Characteristics of Lexical Bundles

Structural patterns of the lexical bundles employed by the NS and the NNS in theoretical linguistics are addressed in this section. Biber et al.'s (1999) taxonomy was slightly modified by referring to Salazar (2011). Table 7 presents the structural types of lexical bundles from the NS corpus.

Table 7. Frequency of structural categories of lexical bundles in the NS corpus

Structure		Types	%	Tokens	%
NP-based	NP - with "of" fragment	21	19.6	396	18
	NP - without "of" fragment	3	2.8	42	1.9
NOUN PHRASE IN TOTAL		24	22.4	438	19.9
PP-based	PP - with "of" fragment	19	17.8	418	19
	PP - without "of" fragment	24	22.4	606	27.6
PREPOSITIONAL PHRASE IN TOTAL		43	40.2	1024	46.6
VP-based	1) Copula *be* + noun/adjP	3	2.8	44	2
	2) VP with active verb	0	0	0	0
	3) Anticipatory *it* + VP/adjP (+ complement clause)	9	8.4	184	8.4
	4) Passive verb + PP fragment	5	4.7	67	3.1
	5) (VP +) *that*-clause fragment	5	4.7	129	5.9
	6) (V/Adj +) *to*-clause fragment	8	7.5	144	6.6
	7) VP with personal pronoun *I/we*	5	4.7	77	3.5

	8) Others	1	0.9	16	0.7
VERBAL PHRASE IN TOTAL		36	33.7	661	30.2
Adjectival phrase		1	0.9	12	0.5
Adverbial phrase		2	1.9	29	1.3
Others		1	0.9	33	1.5
TOTAL		107	100	2197	100

Table 7 exhibits that the NS used PP-based bundles the most. The second most commonly used pattern was VP-based, followed by NP-based bundles. Adverbial-clauses, adjectival phrases, and others were rarely seen in their texts.

As for the NNS corpus, Table 8 provides a summary of the lexical bundles used by the NNS in structural types.

Table 8. Frequency of structural categories of lexical bundles in the NNS corpus

Structure		Types	%	Tokens	%
NP-based	NP – with "of" fragment	39	14.5	389	13.5
	NP – without "of" fragment	19	7	174	6
NOUN PHRASE IN TOTAL		58	21.5	563	19.5
PP-based	PP – with "of" fragment	33	12.2	402	13.9
	PP – without "of" fragment	37	13.7	659	22.8
PREPOSITIONAL PHRASE IN TOTAL		70	25.9	1061	36.7
VP-based	1) Copula *be* + NP/adjP	18	6.7	140	4.8
	2) VP with active verb	9	3.3	70	2.4
	3) Anticipatory *it* + VP/adjP (+ complement clause)	14	5.2	103	3.6
	4) Passive verb + PP fragment	30	11.1	268	9.3
	5) (VP +) *that*-clause fragment	14	5.2	150	5.2
	6) (VP/Adj +) *to*-clause fragment	11	4.1	90	3.1
	7) VP with personal pronoun *I/We*	10	3.7	120	4.2
	8) Others	5	1.9	32	1.1
VERBAL PHRASE IN TOTAL		111	41.2	973	33.7
Adjectival-phrase		7	2.6	56	1.9
Adverbial-clause		22	8.1	206	7.1
Others		2	0.7	31	1.1
TOTAL		270	100	2890	100

From Table 8, we can see a similar pattern for the NNS. PP-based bundles were the most common structures in token, but came second to VP-based bundles in type. The VP-based bundles were most frequently found in type but were second-most common in token. NP-based bundles followed third both in type and token. Adverbial-clause fragments were listed next. Such inconsistencies between type and token in PP-based bundles of the NNS corpus suggest that certain bundles were enormously adopted, unlike in the NS corpus. Furthermore, the fact that the NNS used the adverbial-clause bundles five times more than the NS implies that the NNS had a tendency to rely more heavily on metadiscursive bundles, which will be discussed later.

Table 9 compares the structural types of the NS and the NNS corpora in theoretical linguistics by absolute frequencies, or tokens, along with relative frequencies per 100,000 words. The relative frequencies are used to compare the results of the two corpora with the different corpora sizes. The relative overuse and underuse by the two groups are measured by the relative frequencies.

Table 9. Frequency of structural categories of lexical bundles in the NS and the NNS corpora

Structure		NS		NNS	
		Raw freq.	Rel. freq.	Raw freq.	Rel. freq.
NP-based	NP – with "of" fragment	396	66.34	389	124.54
	NP – without "of" fragment	42	7.04	174	55.71
NOUN PHRASE IN TOTAL		438	73.38	563	180.24
PP-based	PP – with "of" fragment	418	70.03	402	128.70
	PP – without "of" fragment	606	101.52	659	210.98
PREPOSITIONAL PHRASE IN TOTAL		1024	171.55	1061	339.67
VP-based	1) Copula *be* + NP/adjP	44	7.37	140	44.82
	2) VP with active verb	0	0	70	22.41
	3) Anticipatory *it* + VP/adjP (+ complement clause)	184	30.83	103	32.97
	4) Passive verb + PP fragment	67	11.22	268	85.80
	5) (VP +) *that*-clause fragment	129	21.61	150	48.02

	6) (V/Adj +) *to*-clause fragment	144	24.12	90	28.81
	7) VP with personal pronoun *I/We*	77	12.90	120	38.42
	8) Others	16	2.68	32	10.24
VERBAL PHRASE IN TOTAL		661	110.74	909	291.01
Adjectival-phrase		12	2.01	56	17.93
Adverbial-clause		29	4.86	206	65.95
Others		33	5.53	31	9.92
TOTAL		2197	368.07	2890	925.22

Table 9 illustrates bundles' structural differences between the two groups. Notably, the NNS preceded the NS in all the categories. They heavily relied on the use of noun phrases without *of*-fragment structures, prepositional phrases without *of*-fragments, and copula-*be* followed by noun or adjectival phrases. NNS also notably used passive verb structures, *that*-clause fragments, verbal phrases with personal pronoun *I*/*we*, and adverbial clauses.

4.2.1 NP-based Bundles

The proportion of NP-based bundles was approximately the same between NS and NNS corpora. The NS employed 22.4% of NP-based bundles in type and 19.9% in token, while the NNS adopted 21.5% of NP-based bundles in type and 19.5% in token. When calculating the relative frequencies, underused bundles by the NS were observed both in the NP-based bundles with and without *of*-phrase fragments. Bundles without *of*-phrase fragments were strikingly underused. Although some of the NP-based bundles employed by both groups looked similar, they diverged in how certain bundles were used. Table 10 lists all NP-based bundles from the NS and the NNS corpora.

Table 10. NP-based bundles

	NS	NNS
NP+ *of* structures	a special case of a subset of the	a change of state a result of the

	the beginning of the the content of the the effects of the the full range of the interpretation of the the left edge of the meaning of the the nature of the the presence of a the position of the the probability of the the question of whether the rest of the the right edge of the role of the the set of all the source of the the structure of the the use of the	a state of affairs and the degree of goal of the present goal of this paper no main effect of one of the most presence or absence of significant main effect of significant main effects of the analysis of the the content of the the duration of the the framework of a the frequency of the the influence of the the left edge of the length of the the location of the the majority of cases the meaning of the the nature of the the perspective of the the position of the the presence of the the rest of the the result of the the results of the the right edge of the same type of the size of the the source of the the target of the the two types of the use of a the use of the
NP structures	the degree to which the same way as the way in which	a crucial role in a significant main effect a situation in which a unified account for all else being equal an increase in the condition than in the difference between the two further support for the no support for the supporting evidence for the the case with the the difference between the the difference in the the extent to which the present study also the present study however the present study therefore the relationship between the

NP + *of* structures were more popular in both groups. They were primarily used to mark properties, existence, locations, groupings, to illustrate actions, to indicate degree and proportions, and to denote qualities. NP structures were mostly employed to compare differences and to describe the methods and results, and to denote degree. On the other hand, the NNS distinctively employed certain NP-based bundles as displayed in Table 11.

Table 11. NP-based bundles exclusively adopted by the NNS

Reported results	a result of the, the result of the, the results of the, the results showed that, significant main effect of, significant main effects of, a main effect of, no main effect of, of the effect of
Showed comparisons	the relationship between the, the difference between the, difference between the two, the difference in the
Demonstrated metadiscursive knowledge	the present study however, goal of this paper, goal of the present, the present study therefore, the present study also

(1) The exact location of the sensor coils on the tongue body varied from speaker to speaker, depending on *the size of the* tongue, but it was placed on the rearmost point when the tongue was pulled out, which was about 4.5-5.5 cm from the tongue tip. (NNS012)

(2) Although the participants read the sentences binocularly, only *the position of the* right eye was monitored. During reading, the participants were supported by a chin/head rest to reduce their head movement. (NNS024)

(3) The clear speech mode induced a decrease in overall speaking rate (i.e., fewer syllables per second) and *an increase in the* number of prosodic phrases (i.e., more IPs per sentence). (NNS016)

Bundles in the examples were exclusively employed by NNS. Such tendency was more likely due to the experimental nature of the studies included in the NNS corpus than idiosyncrasies of some writers.

4.2.2 Verb Structures

Verb constructions were more often adopted by NNS than by NS in type, with only a small difference in token. Delving into the subcategories, verb phrase structures with personal pronoun *I/we*, *that*-clause fragments, passive verbs with prepositional-phrase fragments, and copula-*be* constructions made the differences between NS and NNS. Comparatively, there were no notable differences in the anticipatory-*it* structures along with *to*-clause fragments. Table 12 lists all verb structures adopted by the NS and the NNS.

Table 12. Verb structures

	NS	NNS
Copula *be* + NP/adjP	is a matter of this is not the this is the case	however there was no is a set of is attributable to the is based on the is due to the is part of the is responsible for the is the same as is used with the may be due to the present study is there was a significant there was also a there was no significant this is because the this is not the to be associated with was significant only in
VP with active verb		belong to the same does not have a do not have a do not have to has a three way participated in the experiment play a role in shed light on the the present study has
Anticipatory *it* + VP/adjP (+ complement clause)	it is clear that it is easy to it is important to it is necessary to it is not a it is not clear it is possible that it is possible to	it will be shown it is possible that it is predicted that it is notable that it has been observed it is expected that it seems that the it is hard to

	it may be that	it is possible to it is important to it is not clear it is difficult to it has been shown it has been suggested
Passive verb + PP fragment	be thought of as can be interpreted as can be seen in is based on the this is illustrated in	are given in the are shown in table are summarized in table be accounted for by be attributed to the be referred to as be related to the be thought of as can also be used can be accounted for can be explained by can be interpreted as can be separated by can be understood as can be used with cannot be attributed to cannot be separated by cannot be used in examples are provided in is attached to the is followed by a is motivated by the is supported by the it should be noted prediction is borne out speakers were asked to this is illustrated in was observed in the were divided into two will be discussed in
(VP +) *that*-clause fragment	that it is not that there is a that there is no that there is some the fact that the	not the case that that it is not that the degree of that the present study that there is a that there is no the assumption that the the case that the the claim that the the fact that the the possibility that the the results showed that the view that the this suggests that the
(V/adj +) *to*-clause fragment	appears to be a can be used to does not seem to seems to be a to account for the	can be used to is expected to be is more likely to present study is to to account for the

	to be able to to be the case turns out to be	to do with the to examine whether the turned out to be used to refer to was found to be
VP with personal pronoun *I/We*	as we have seen as we will see i will argue that we have already seen we have the following	as we will see i argue that the i have argued that i have shown that i propose that the we assume that the we expect that the we have seen that we predict that the we propose that the
Others	may or may not	but they do not discussed in the previous does not have to let us first consider let us now consider

4.2.2.1 VPs with Personal Pronoun *I/We*

An examination of the concordances revealed that NS showed variations in the subject selection and the use of modifiers. *We predict that the* is a commonly used bundle in NNS corpus to construct the writer's argument. Although it was not found in the NS corpus, a search for expressions that incorporated *predict that* resulted in 28 instances of the same use as the target bundle. However, a striking difference was in the selection of the subject. While the NNS mainly stuck to the bundle *we predict that the*, the NS exhibited a wide range of use in selecting the subject as seen below:

(4) If we remove the lexical item [author:+], then we still have an analysis *that will predict that* we were and we was are both possible (using Lis (b) and (c) only). (NS046)

(5) This will allow the form that to appear with plural nouns, since the demonstrative is unspecified for agreement. *It will also predict that* verbal –s with such an DP should be possible, since we can pick T specified with [pronominal:–]. (NS042)

In (4) and (5), the authors displayed varied uses of synonymous expressions by which they concealed their being subjects. *An analysis* and *it* instead took their places.

Furthermore, NS demonstrated a highly varied use of modifiers such as adverbs before the verb *predict*.

(6) *We therefore correctly predict that* RAs should pass neither the tests for having a maximal element nor the tests for having a minimal element. (NS031)

(7) Thus *we automatically predict that*, for instance, (43b) has an internal reading that entails that there is a unique book such that proper subparts of the non-atomic entity consisting of John and Bill each read that book, as desired. (NS027)

In (6), the author wove the strand of logic by inserting *therefore* and showed confidence in the argument by placing *correctly* before the verb *predict*. In (7), the author indicated that he reached a logical conclusion by stating *automatically* before *predict*. Likewise, such flexibilities played a role in the lower number of retrieved lexical bundles from the NS corpus. By contrast, NNS' lack of variation in tense, subject, and modifier contributed to the higher number of lexical bundles produced.

4.2.2.2 Passive Verb Structures

Passive verb constructions were the most significant in overall structural differences. While the NS employed only 5 types of the bundles, the NNS exhibited 31. A decisive factor in the NS' substantial underuse of passive verb combinations over the NNS was the use of the active voice. *Can be accounted for* and *prediction is borne out* were 2 bundles that existed only in the NNS corpus. A search for the synonyms, however, revealed that whereas the NNS did not use *we can account for*, the NS used it 3 times. In the same vein, as a substitute for *prediction is borne out*,

the active form *confirm* was used 7 times along with 3 instances of *we can show* in the NS corpus. By contrast, the NNS used *confirm* only twice alternatively. The results imply that the NNS prefer the passive to the active form in academic writing.

4.2.2.3 Copula-*be* Constructions

Copula-*be* constructions were other bundles in which the NNS exceeded the NS in use. The NNS had 140 tokens while the NS used 44. The lexical bundle in the NNS corpus that recorded the highest hit was *there was a significant*, with 13 instances. One notable finding was in the absence of the bundle in the NS corpus. However, when *significant* was typed in the search box of the NS corpus, 177 hits were found. Among them, 73 instances were of the same usage.

(8) Random slopes for Word Duration reached significance (sd¼0.035) and entered into a *significant* correlation with the by-subjects random intercepts (r¼0.793), indicating that slower subjects had a larger effect of Word Duration. (NS005)

(9) The HWD experiment shows an extraordinary number of main effects (HWD Tables 7, 8) and a number of *significant* interactions. (NS021)

(10) Most importantly, the interaction effect remained *significant*: violating an unnatural constraint resulted in a smaller reduction in participant ratings than violating a natural constraint. (NS009)

(11) The onset factor is *significant* as a predictor of initial vs. final stress ($F(2) = 12.7$, $p < .0001$), with 43% initial stress for ø, 60% for C, and 79% for CC. (NS015)

As seen in the examples, unlike the NNS who preferred the fixed phrase *there was a significant*, the NS exhibited a much wider and less restricted use of similar expressions. *Random slopes* (8), *the HWD experiment* (9), *the interaction effect* (10), and *the onset factor* (11) substituted the subjects where the NNS would have simply used the existential-*there*. Likewise, the bundle *is attributable to the* was not observed in the NS corpus, but 8 instances of "attributable" were found in it. The NS substitutes for these included modal verbs together with a modifier before *to* or a reflexive pronoun, again displaying variations that were not found in the NNS corpus.

(12) …, because any differences *could be attributable* either *to* inherent differences in the processes themselves or *to* differences between affricates in isolated words on the one hand and complete sentences on the other. (NS004)

(13) …, but the upshot is that for the current study any difference in the behaviour of postalveolar vs. alveolar affricates *could be attributable* not just *to* place, but also *to* number of syllables, vowel length or position relative to stress (which is always initial in Hungarian), or some combination of these factors. (NS004)

(14) …; this proposal has led to various later interpretations (Rice 1992, Kager 1993, Everett 2003, Revithiadou 2004), and is potentially related to a preference for weak positions following stresses (Hung 1994, van de Vijver 1998, Hyde 2007a), *itself attributable to the* typical need for a peak followed by a trough to realize HL intonation (Bolinger 1962, Hyman 1977). (NS008)

In (12), the modal verb *could* was used to indicate the degree of certainty. In (12), *either* is placed before *to* in order to show a larger pool of possibility. In (13), *not just* preceded *to* in

order to emphasize the importance of the second cause. In (14), the author took *this proposal* as the subject in a form of reflexive pronoun. Such varieties also contributed to reducing the retrieved bundles from the NS texts.

4.2.3 Prepositional-phrase Fragments

Table 13 displays prepositional-phrase fragments used by the NS and the NNS.

Table 13. Prepositional-phrase fragments

	NS	NNS
PP + *of* structures	along the lines of as part of the at the end of for the sake of in a variety of in terms of a in the absence of in the analysis of in the case of in the context of in the form of in the sense of of the examples in of the same type of the theory of on the basis of to the left of to the right of to the set of	about the nature of as a function of as a result of as one of the at the beginning of at the edges of at the end of at the time of for the occurrence of for the purpose of for the sake of in favor of the in terms of the in the case of in the context of in the framework of in the interpretation of in the majority of in the sense of in the state of in the use of in the vicinity of of the effect of of the present study on any of the on the basis of on the nature of on the type of regardless of whether the to that of the
PP structures	at the left edge at the level of at the right edge at the same time by the fact that from the fact that in a way that	as a result the as to whether the at least in part at the right edge at the same time by the fact that contrary to the facts

	in addition to the in other words the in terms of the in the next section in the previous section in the same way in the sense that in this case the in this section i in this section we on the assumption that on the one hand on the other hand such as those in to the extent that to the right edge with respect to the	depending on the context despite the fact that due to the fact for example in the from the fact that in a situation where in addition to the in contrast to the in line with the in other words it in other words the in such a case in such a way in the current study in the examples in in the following section in the next section in the present study in the previous section in this paper i in this paper we in this section i in this section we on the one hand on the other hand such as that in to the extent that under the assumption that with respect to the with the fact that

The NNS showed a significant overuse of prepositional-phrase bundles over the NS. PP + *of* structures described methods, presented results, denoted degrees and amounts, and signified logical relations. Notably, PP structures were mostly metadiscursive expressions for both groups of researchers. Upon a search of concordances, several differences in usage emerged. *On the other hand*, the most common pattern from both NS and NNS corpora, showed marked differences in usage. It on the one hand shared the meanings of transition and addition (15, 16) and comparison and contrast (17, 18), but on the other hand had distinct usages of its own (19, 20, 21).

(15) In a grammar where a harmonically bounded candidate C is the only possible output, higher ranked constraints must eliminate the other candidates. But in that case, C is not in fact harmonically bounded. Consequently, generation of

harmonically bounded outputs entails generation of their harmonic bounders too. *On the other hand*, constraints can be added to reduce the set of viable harmonically bounded forms. (NS001)

(16) If age-grading is responsible for the age-conditioned variation, the two lines will coincide with each other as shown in Fig. 2(a). So, we expect more conservative pattern in the speech of 81-year-old speaker in 2005 than in the speech of the two 1935 speakers.[2] *On the other hand*, if the synchronic variation in Present Day Korean is due to sound change in progress, we expect that the 41-year-old speaker in 1935, who is a whole generation ahead of the 81-year-old speaker of 2005 in real time, to produce a more conservative speech pattern than the 81-year-old speaker as schematically shown in Fig. 2(b) and (c). (NNS014)

(17) As expected, the weights for the unnatural constraints fell slightly when relevant exceptions were added to the training set. The natural constraints, *on the other hand*, received higher weights when the affixed forms were added, probably as a result of the larger set of data for which they could "prove their worth" by remaining exceptionless. (NS009)

(18) Accent did not interact with the Listener group, either, suggesting that its null effect was consistent across listener groups. CD, *on the other hand*, yielded a significant main effect ($F\ [1,57] = 7.25$, $p<0.01$), but interacted with Boundary and Accent as discussed above. (NNS009)

The NS uniquely used the bundle in the middle of a sentence, which was not prevalent among the NNS' corpus. Their intention was for the two different personal affective judgments (19) or when leading the reader through the text to the author's arguments (20). In (20), the author considers two possibilities in reasoning and employs the structure after having dealt with the first.

(19) That there are so few results is on the one hand unsurprising, considering that Patterson & Connine (2001) find a flapping rate of about 93.9% in disyllables, but *on the other hand* worrisome. (NS001)

(20) If movement applies, the order 'WhP > V > X' will be formed, in contradiction of the earlier order. Suppose *on the other hand* that before Spell Out applies to VP, the Wh-phrase has moved to the (left) edge of VP as in (62): (NS030)

In contrast, the NNS also used the pattern when the context did not include a change of standpoint. In this case, the adversative conjunction *however* looked more appropriate.

(21) The reader could interpret a negation as a DN first and then, at the clarification clause, reanalyze it into an MN, taking more processing time for the MN. *On the other hand*, he or she could decide on the target of negation at the clarification clause by using the contextual information provided by the clause. (NNS020)

Another notable difference was in the presence of metadiscursive bundles in the NNS corpus. Among the prepositional-phrase bundles without *of*-phrase fragments, metadiscursive bundles amounted to 22.9% in the NNS corpus. While such patterns abounded, the NNS lacked in strategies to unfold what was just mentioned. On the other hand, the NS had a wider means of developing the text by further explicating what had been presented.

(22) Both require the marked word or the corresponding concept to be not-given, *in the sense that* some alternative elements of the same type must be contextually available, while everything else must be given, in the sense of accommodatable as a property of the alternative set in question. (NS020)

(23) An alternative account that I will sketch here, but ultimately reject, involves the observation noted earlier that the two classes of adjectives contrast *in a*

way that could reasonably be related to the Stage/Individual distinction of Carlson (1977) and others. (NS031)

In (22), the metadiscursive device *in the sense that* serves to qualify the requirement in the main clause. Likewise, in (23), *in a way that* is used to explain the contrast that precedes the lexical bundle. Such a strategy seems useful, for it can give richer understanding of a presented statement.

4.2.4 Summary

Structural characteristics of lexical bundles used by NS and NNS linguists were explored in this section. Both groups' structural pattern of use was generally similar, with the PP-based the largest, the VP-based the second, and NP-based bundles the least. This sequence marks a distinction from the literature, which noted that NS preferred the NP-based bundles, whereas NNS opted for the VP-based (Chen & Baker, 2010; Biber et al., 1999). NNS' dependence on particular bundles lends credence to Salazar's claim that the NNS have a tendency to overuse certain familiar bundles and underuse relatively unknown ones (Salazar, 2011). Such a tendency was termed "lexical teddy bears" (Hasselgren, 1994: 237) and has been attested by researchers (Ädel & Erman, 2012; Granger, 1998; Hasselgren, 1994). Furthermore, the NNS preferred passive forms to active voice in academic writing. This strategy can minimize the personal role of the author and thus disguise authorial identity in the interpretation of data (Hyland, 2008b). In other words, NNS were more likely to submit to the academic conventions of author anonymity. In addition, diverse usages of the same bundle were observed between the two different groups.

4.3 Functional Characteristics of Lexical Bundles

In this section, functional characteristics of the identified lexical bundles are discussed. Table 14 shows the lexical bundles from the NS corpus, functionally categorized with their frequencies of both type and token.

Table 14. Functional classification of lexical bundles in the NS corpus

Function	Types	%	Tokens	%
Research-oriented bundles	35	27.7	688	27.3
Location	8	6.3	159	6.3
Procedure	6	4.8	112	4.5
Quantification	1	0.8	15	0.6
Description	12	9.5	245	9.7
Grouping	8	6.3	157	6.2
Text-oriented bundles	69	54.8	1466	58.1
Additive	3	2.4	98	3.9
Comparative	6	4.8	146	5.8
Inferential	20	15.9	338	13.4
Causative	3	2.4	52	2
Structuring	10	7.9	161	6.4
Framing	26	20.6	625	24.8
Citation	0	0	0	0
Objective	1	0.8	46	1.8
Participant-oriented bundles	22	17.5	368	14.6
Stance	15	11.9	249	9.9
Engagement	7	5.6	119	4.7
Total	126	100	2522	100

In the NS corpus, text-oriented bundles occupied the largest proportion, followed by research-oriented bundles, with the participant-oriented the least. Of the text-oriented bundles, the NS preferred framing bundles over others and favored inferential bundles next to the framing bundles. They also showed a tendency to reveal stance than engaging with their readers by use of lexical bundles.

Table 15 provides a brief overview of the functional classification of the lexical bundles in the NNS corpus by type and token.

Table 15. Functional classification of lexical bundles in the NNS corpus

Function	Types	%	Tokens	%
Research-oriented bundles	66	19.9	588	17
Location	11	3.3	112	3.2

Procedure	12	3.6	89	2.6
Quantification	7	2.1	73	2.1
Description	27	8.2	247	7.2
Grouping	9	2.7	67	1.9
Text-oriented bundles	197	59.5	2203	63.9
Additive	3	0.9	43	1.2
Comparative	18	5.4	317	9.2
Inferential	55	16.6	465	13.5
Causative	29	8.8	295	8.6
Structuring	41	12.4	478	13.9
Framing	37	11.2	511	14.8
Citation	10	3	66	1.9
Objective	4	1.2	28	0.8
Participant-oriented bundles	68	20.6	660	19.1
Stance	48	14.5	452	13.1
Engagement	20	6.1	208	6
Total	331	100	3451	100

A similar trend was observed in the NNS linguists, who favored text-oriented bundles. Research-oriented bundles and participant-oriented bundles followed next. As can be observed from Tables 14 and 15, text-oriented bundles were most commonly adopted by both the NS and the NNS. This preponderance of text-oriented bundles is a clear indicator of highly-advanced writers (Hyland, 2008a). Possibilities are that the control of genre, discipline, and expertise in data collection were crucial factors in explaining the similarity in the NS and the NNS groups in the present study. However, the two groups also exhibited differences. The NS displayed a more balanced pattern among the three types of bundles, whereas text-oriented bundles were the highest with the others at a lower number in the NNS. This difference originated from a larger number of research-oriented bundles in NS corpus.

Table 16 compares the functional types of the NS and the NNS corpora in theoretical linguistics by absolute frequencies, together with relative frequencies.

Table 16. Frequency of functional categories of lexical bundles in the NS and the NNS corpora

Function	NS		NNS	
	Raw freq.	Rel. freq.	Raw freq.	Rel. freq.
Research-oriented bundles	688	115.26	588	188.25
Location	159	26.64	112	35.86
Procedure	112	18.76	89	28.49
Quantification	15	2.51	73	23.37
Description	245	41.05	247	79.08
Grouping	157	26.30	67	21.45

Text-oriented bundles	1466	245.60	2203	705.28
Additive	98	16.42	43	13.77
Comparative	146	24.46	317	101.49
Inferential	338	56.63	465	148.87
Causative	52	8.71	295	94.44
Structuring	161	26.97	478	153.03
Framing	625	104.71	511	163.59
Citation	0	0	66	21.13
Objective	46	7.71	28	8.96
Participant-oriented bundles	368	61.65	660	211.30
Stance	249	41.72	452	144.71
Engagement	119	19.94	208	66.59
Total	2522	422.52	3451	1104.82

From Table 16, it is shown that the NNS enormously exceeded the NS in the use of all but grouping, additive, and objective bundles. Grouping and additive bundles are the only ones the NNS surpassed the NS. They employed objective bundles slightly more often than the NS. Quantification and causative bundles are the particular areas in which the NNS displayed frequent use, to more than ten times of the NS in relative frequencies. Such inclinations were explained by the experimental nature of the NNS' selected articles. They also demonstrated heavy uses in both types of participant-oriented bundles, stance and engagement.

4.3.1 Research-oriented Bundles

In both groups, research-oriented bundles were second most common, next to text-oriented bundles. Table 17 lists all research-oriented bundles adopted by the NS and the NNS.

Table 17. Research-oriented bundles

	NS	NNS
Location	at the end of at the left edge at the right edge the left edge of the right edge of to the left of to the right edge to the right of	at the beginning of at the edges of at the end of at the right edge first and the second in the vicinity of the first and second the left edge of the location of the the position of the the right edge of
Procedure	in the analysis of	all else being equal

	that there is no that there is some the beginning of the the content of the the use of the	an analysis of the at the time of belong to the same can be separated by for the purpose of over the course of participated in the experiment speakers were asked to the analysis of the the target of the were divided into two
Quantification	the probability of the	about cm from the an increase in the the difference in the the duration of the the frequency of the the length of the the size of the
Description	at the level of is a matter of that there is a the degree to which the meaning of the the nature of the the position of the the presence of a the role of the the structure of the to be able to to the extent that	a change of state a state of affairs about the nature of and the degree of at least in part be referred to as can be used with do not have a does not have a has a three way in the state of in the use of it has been suggested on the degree of on the nature of on the type of presence or absence of that the degree of the content of the the extent to which the meaning of the the nature of the the presence of the the use of a the use of the to the extent that used to refer to
Grouping	a subset of the as part of the in a variety of of the same type the full range of the rest of the the set of all to the set of	as one of the in the majority of is a set of on any of the one of the most the majority of cases the rest of the the same type of the two types of

In both the NS and the NNS corpora, the most frequently occurring type of research-oriented bundles was description bundles. The majority of them was in "noun phrase + *of*" structures and devoted to describing research objects or contexts. As discussed earlier, a significant proportion of research-oriented description bundles suggests the unbiased and objective nature of research articles (Hyland, 2008b). Among the research-oriented bundles, only grouping bundles was the category where the NS overused in comparison with the NNS. Although there was not a big difference between the groups, one noteworthy finding was in the NS' unique use of certain bundles.

Of the retrieved data, *a subset of the* was not found in the NNS list. An examination of the concordances attested that *a subset of the* (31) was exchangeable with *a part of the*.

(24) The Variably Strict account's explanation of this phenomenon is straight forward: the closest possibilities where Bob goes to the parade and gets stuck behind someone tall are not *a subset of the* closest possibilities where Bob goes to the parade. (NS034)

In addition, the NS adopted *the full range of* to denote the full scope, or all.

(25) If all sprouting could always be analyzed as anaphora to a full clause, it would be easy to generalize Jäger's (2001, 2005) analysis to cover sprouting. However, this strategy is unlikely to generalize to *the full range of* sprouting examples. (NS029)

(26) While the present account explicitly recognizes a family of related constructions in order to account for the evident semantic variability, it may seem that other approaches yield greater generalizations. Few accounts, however, have even attempted to account for *the full range of* data discussed here.(NS039)

Furthermore, *in a variety of* was exclusively found in the NS corpus, instead of *many*.

(27) Beyond their role in helping to establish stress patterns, feet play an important part *in a variety of* phenomena. (NS006)

(28) For external Merge, these go beyond argument structure associated with substantive categories, and they presumably include functional categories of the kind coming to light *in a variety of* inquiries, notably the very fruitful cartography projects (Cinque 2002, Rizzi 2004, Belletti 2004). (NS041)

The identified bundles on the list were too small in number to fully account for the occurrences. However, it revealed that the NS were more likely to use a wide range of bundles that denoted grouped individuals or its parts. Location bundles were mostly used to describe research objects as in phonetics/phonology and syntax, such as *the left edge of*, *to the right edge*, and *at the end of*. The less frequent procedure bundles served to report research procedures and analysis, usually found in the methods section. Quantification bundles were the least common bundles, with 1 type found in the NS corpus and seven in the NNS corpus. They functioned to indicate measures, quantities, and proportions, as in *the duration of the*, *the size of the*, and *the probability of the*.

4.3.2 Text-oriented Bundles

Text-oriented bundles accounted for the largest proportion of all functional categories, occupying more than the half of the NS corpus and two-thirds of the NNS corpus. Table 18 lists text-oriented bundles by NS and NNS.

Table 18. Text-oriented bundles

	NS	NNS
Additive	as well as the at the same time in addition to the	as well as in as well as the in addition to the
Comparative	along the lines of in the same way on the one hand on the other hand the same as the	as compared to the as in the case as opposed to the as was the case condition than in the

	the same way as	contrary to the facts difference between the two different from that of in contrast to the in line with the is the same as on the one hand on the other hand similar to that of the difference between the the relationship between the the same way as to that of the
Inferential	appears to be a be thought of as can be interpreted as can be used to does not seem to i will argue that it is clear that it is easy to it is necessary to it is not a it is not clear it is possible that it is possible to it may be that may or may not seems to be a that it is not the interpretation of the this is not the turns out to be	be accounted for by be attributed to the be thought of as can also be used can be explained by can be interpreted as can be used to cannot be separated by cannot be used in further support for the i argue that the i propose that the is assumed to be is attributable to the is expected to be is more likely to is motivated by the it is expected that it is hard to it is possible that it is possible to it is predicted that it seems that the prediction is borne out that it is not that there is a that there is no the possibility that the the results showed that there was a significant there was no significant this suggests that the was found to be was observed in the we expect that the we predict that the
Causative	for the sake of the effects of the the source of the	a crucial role in a result of the a significant main effect a unified account for as a result of as a result the be accounted for by

		be attributed to the by the fact that can be accounted for cannot be attributed to due to the fact for the occurrence of is attributable to the is due to the is responsible for the is supported by the may be due to no main effect of of the effect of play a role in significant main effect of significant main effects of the influence of the the result of the the results of the the source of the this is because the
Structuring	can be seen in in the next section in the previous section in this section i in this section we of the examples in such as those in this is illustrated in we have already seen we have the following	are given in the are summarized in table as described in the as illustrated in the as in the example as shown in fig as shown in figure as shown in table as shown in the for example in the in the current study in the examples in in the following section in the next section in the present study in the previous section in this paper i in this paper we in this section i in this section we is followed by a it will be shown let us first consider of the present study such as that in that the present study the present study has the present study however the present study is this is illustrated in will be discussed in
Framing	as far as i by the fact that from the fact that if and only if	a situation in which as a function of as far as i as far as the

	in a way that in other words the in terms of a in terms of the in the absence of in the case of in the context of in the form of in the sense of in the sense that in this case the is based on the of the theory of on the assumption that on the basis of the fact that the the question of whether the way in which this is the case to be the case with respect to the	as to whether the at the same time despite the fact that from the fact that in a situation where in favor of the in other words the in such a case in such a way in terms of the in the case of in the context of in the framework of in the interpretation of in the sense of is based on the not the case that on the basis of regardless of whether the the assumption that the the case that the the case with the the claim that the the fact that the to do with the under the assumption that when there is no with respect to the with the fact that
Citation		as pointed out by can be understood as do not have to in other words it is attached to the is part of the is used with the it has been observed it has been shown to be associated with
Objective	to account for the	in order to examine to account for the to examine whether the whether and how the

4.3.2.1 Framing Bundles

The most frequently appearing bundles in both corpora were framing in token. Framing bundles served to elaborate arguments by specifying conditions and highlighting limitations. The present study found the NNS scholars' tendency to overuse bundles. Nonetheless, some bundles existed that were not attested in the NNS corpus; *In a way that* (29, 30) and *in the sense that* (31) were

found exclusively in the NS corpus with the respective occurrences 30 and 29. These bundles all shared a common role of qualifying a proposition or a statement. Such bundles were not found in the NNS corpus.

(29) Like other properties of our learner, our proposals concerning the search space and heuristics constitute a theoretical claim about language learning. To be sure, they are also motivated by issues of implementation — but not, we think, *in a way that* sacrifices realism with respect to the human learner. (NS010)

(30) An alternative account that I will sketch here, but ultimately reject, involves the observation noted earlier that the two classes of adjectives contrast *in a way that* could reasonably be related to the Stage/Individual distinction of Carlson (1977) and others. (NS052)

(31) The distribution of coda [h] in Huariapano is thus 'rhythmic', *in the sense that* it picks out every other syllable in the word. (NS018)

In the examples, the expressions served to modify a preceding proposition. In other words, they set a confinement to the already mentioned statement. This use was exclusive to the NS authors, and they can thus be called "native bundles".

4.3.2.2 Inferential Bundles

The second most common ones among the text-oriented bundles were different in the NS and the NNS. Inferential bundles were second most common among the NS both in type and token, whereas they were listed as the largest type with the third largest frequency among the NNS. Overall, inferential bundles were more enjoyed by the NS, while the NNS preferred structuring bundles. A contextual analysis revealed that the major difference in inferential bundles between the two groups was in the presence of anticipatory-*it* structures and verbs with *I/we*. Whereas the inferential anticipatory-*it* structures accounted for 8.4% of the overall tokens in the NS

corpus, *I/we* verbs occupied only 4.7%. On the other hand, the NNS adopted more *I/we* verbs than anticipatory-*it* structures as seen in the respective 4.2% and 3.6%. The higher use of *I/we* verbs by the NNS showed a marked contrast with the findings of Salazar (2011). By referring to Vassileva's (1998) taxonomy of functions of *I* in academic writing, none of the *I/we* verbs in the NS corpus conveyed personal view, while 56.7% of the bundles employed by NNS revealed authorial view. This implies that the NS had other means to express their opinions than by the *I/we* form. Turning to the anticipatory-*it* structures and some other patterns including *turns out to be, can be interpreted as, appears to be a, does not seem to,* and *seems to be a* provides a clue about this point. Rather than conveying arguments by the use of the subject *I/we*, the NS relied more on indirect devices such as stance markers to present their reasoning based on the demonstrated data.

(32) We have two conjectures for the presence of these puzzling constraints. First, *it is possible that* they are valid, by which we mean that had it been possible to carry out experiments with Wargamay speakers of the kind Scholes performed, it would have emerged that test forms violating these constraints were rated low. (NS010)

(33) Alternatively, or perhaps additionally, *it may be that* there is featurally [pronominal:–], perhaps because it has a locative complement to its D or because it originates as part of the associate (as in Kayne 2008). (NS044)

(34) The fact that embedded Thetic Subjects cannot undergo SOR appears to undermine the movement analysis of SOR, since an embedded Thetic Subject is the highest A-specifier of the embedded domain in such cases. *I argue that the* reason Thetic Subjects do not undergo SOR is because the positions of two types of subjects are different. (NNS029)

(35) Given the above pattern, *I propose that the* signification of a sensory observation is lexically encoded in the meaning of *–te*. (NNS027)

In (32) and (33), by adopting the anticipatory-*it* devices, the NS unfold the analyses in more careful patterns, leaving room for readers to argue. This strategy gives to readers a connotation that the presented answer is not definitive and is open to dispute. On the other hand, in (34) and (35), the NNS deliver their messages in more authoritative ways by employing the *I*-form.

Furthermore, there is another factor contributing to the relatively lower proportion of *I/we* verbs in the NS corpus. As discussed earlier, the NS seemed freer in using variations with the verbs – in modal and adverb. As of modals, although *I argue that the* did not make it to the NS list, *I will argue that* was found 15 tokens, *I shall argue that* twice, and *I would argue that* once. Adverb was another factor. Between the subject and the verb in the bundle, *will eventually*, *ultimately*, *will ultimately*, and *then* were inserted and were blocking the construction of the bundle. Additionally, in an instance, it was joined by another verb.

4.3.2.3 Structuring Bundles

Structuring bundles turned out to be the second most frequent in the NNS corpus while they ranked much lower in the NS corpus. As text-reflexive markers, their primary function is "to help organize the stretches of discourse by providing a frame within which new arguments can be both anchored and projected, announcing discourse goals and referring to text stages" (Hyland, 2008b: 16). The structures were used as introductory devices to help the readers gain an overall idea of the whole.

(36) *In this section we* will briefly reformulate what has already been discussed in terms of typed feature structures, describing some implementation choices along the way. (NS024)

(37) *In this paper, we* examine the stop consonant production in an earlier stage of Seoul Korean than has been examined by previous studies. (NNS014)

In (36), the NS researcher offers the sectional frame in which a discourse could be unfolded. Similarly, in (37), the NNS researcher announces the goal of the paper for readers to grasp the general idea of the texts.

Additionally, they referred to other parts of the texts to bring them to the readers' attention.

(38) We believe his own corpus data indicate that the construction is in fact partially productive, in that there are many verbs that occur only once in the 10-million-word corpus with a particular RP. *This is illustrated in* Table 1. (NS050)

(39) A slope less than 1 indicates that there is another factor affecting the timing of the tones; this *will be discussed in* the following section. (NNS001)

In (38) and (39), the structuring bundles work as pointers for the readers to the following contents. Altogether, the structuring bundles in the NS corpus accounted for approximately 6.4% of text-oriented bundles, whereas those in the NNS corpus amounted to 14%. Such a big difference originated in the NNS researchers' heavy use of metadiscursive bundles as has been discussed in the present study.

4.3.2.4 Comparative Bundles

Comparative bundles were also overused by the NNS. These structures were spotted as 6 types in the NS corpus and 18 types in the NNS corpus. Bundles in the NNS corpus such as *as opposed to the, difference between the two, the difference between the, contrary to the facts, in contrast to the, different from that of, as compared to the, to that of the, similar to that of, as in the case* made 84 occurrences in total. The NNS mostly adopted these bundles when comparing different results or examples in analysis.

(40) *As opposed to the* written discourse, the most frequent function, with 70 examples, is depicting a discontinued simple past event. A typical example is provided in (10) below. (NNS021)

(41) As described in (58), it is then expected that the object may move across the resultative predicate, stranding an NQ *in contrast to the* data in (57b). (NNS036)

In (40) and (41), the researchers used the comparison strategy as they unfolded analyses, thus presenting their views more effectively. Additionally, in providing claims or assumptions, the NNS employed comparative bundles to show that the writers' assertions had solid grounds as proven in the previous studies. They used *in line with the* 15 times. Compared to the NS who used only *along the lines of* 13 times, it was an evident overuse, considering that the NNS corpus was of approximately half the corpus size of the NS.

(42) This is *in line with the* view that the locality hypothesis is better characterized as a gradient constraint rather than as an all-or-none constraint as suggested in the literature (e.g,. Byrd & Saltzman, 2003; Cho, 2005; Cho & McQueen, 2005; Byrd, et al., 2006; Cho & Keating, 2009). (NNS007)

In the example, the author employs the bundle in a way to provide supporting basis for the findings of the study. In this way, authors can ensure foundation for the analyses or the arguments made in their own studies.

4.3.2.5 Causative Bundles

Whereas the NS did not prefer causative bundles, the NNS used them as much as they did comparative bundles. The difference can be seen in the contrasting number of types of bundles; while the NS employed 3 types, the NNS adopted 29 types. Most of them were used to present findings and to interpret the results based on the writer's interpretations of the data. The

differences can first be explained in terms of different methodologies. As can be deduced from the NNS' 40 tokens use of the bundles *there was a significant*, *a significant main effect*, and *no main effect of*, which was confirmed by scanning the articles, the NNS corpus consisted of far more articles with experiments. In the NS corpus, experiments were used only in the phonology/phonetics subfield, but semantics and syntax articles also contained experimental data in the NNS corpus. However, to interpret the results as a mere overuse by the NNS is too simple and rather inadequate to represent the compiled data. As it was found earlier in this study, the NS writers' variations in the use of expression had caused the discrepancy. For example, 7 tokens of *is attributable to the* were observed in the NNS corpus. While the same bundle did not exist in the NS corpus, searches for the keywords *attribute*, *attributed*, and *attributable* traced 27, 34, and 8 examples each. The reason for the absence of the target bundle was that the verbs *attribute* and *attributed* were used with a variety of modal verbs including *can be, may be, could be, most likely be, can plausibly be, is necessarily*, and *must be*. Before the adjective *attributable* were *could be, is, are, is arguably*, and *the penalty were* used in combination.

4.3.2.6 Summary

Text-oriented bundles occupied a major proportion in both NS and NNS corpora. The preponderance of text-oriented bundles is obviously a feature of expert writers (Hyland, 2008a), which has also been attested by Salazar (2011). Framing bundles were far more heavily used by NNS. As Hyland (2008a) discovered that highly experienced writers demonstrate a large use of framing bundles, it can be suggested that the control for the same degree of expertise played a role in the present study. Inferential bundles were more likely to be native bundles. The NS tended to express the authors' views using indirect devices, whereas the NNS conveyed their messages by the use of personal pronouns. The fact that NNS were heavy users of structuring bundles substantiates our previous finding that NNS rely on metadiscursive bundles. The NNS' overuse of metadiscursive bundles has been noted elsewhere (Ä del & Erman, 2012; Ä del, 2006). The NNS also took advantage of the use of comparative bundles. They mostly used the bundles when comparing or contrasting examples in analysis or to establish the solid foundation of their

results by aligning them with the literature. Whereas causative bundles were rarely seen in the NS corpus, the NNS abounded with them. The main reason originated in the experimental nature of NNS' articles, but the NS' diversity in language use contributed to smaller bundle numbers.

4.3.3 Participant-oriented Bundles

Participant-oriented bundles are situated in an interpersonal domain and concern writer–reader interaction of the text. They are composed of two kinds – stance and engagement. Stance, which includes writer-oriented features of interaction, lies in an attitudinal dimension and is the means through which writers interrupt to identify their personal authority or camouflage their involvement in the text. Engagement is located in an alignment dimension. Writers use engagement markers to bring the readers into the discourse in anticipation of their reaction to writers' arguments (Hyland, 2005). In this study, the NNS overused participant-oriented bundles as in other patterns, both in stance and engagement markers. Table 19 demonstrates participant-oriented bundles employed by NS and NNS.

Table 19. Participant-oriented bundles

	NS	NNS
Stance	appears to be a can be interpreted as can be used to does not seem to i will argue that it is clear that it is easy to it is important to it is not clear it is possible that it may be that may or may not seems to be a	can also be used can be accounted for can be explained by can be interpreted as can be separated by can be understood as can be used to can be used with i argue that the i have argued that I have shown that i propose that the if we assume that is assumed to be is expected to be is more likely to it is difficult to it is hard to it is important to it is not clear

		it is notable that it is possible that it seems that the it should be noted may be due to reasonable to assume that the possibility that the this suggests that the under the assumption that we assume that the we expect that the we predict that the we propose that the
Engagement	as we have seen as we will see can be seen in it is important to it is necessary to we have already seen we have the following	as can be seen as described in the as discussed in section as illustrated in the as in the case as in the example as reflected in the as shown in fig as shown in figure as shown in table as shown in the as was the case as we will see for example in the if we assume that it is important to it should be noted let us first consider let us now consider we have seen that

4.3.3.1 Stance Bundles

In participant-oriented stance bundles, some interesting discrepancies existed between the NS and the NNS use. Hedges and self-mention strategies are as such. Hedges are devices that signal the writer's intention to reserve whole commitment to a proposition and allow information to be presented as an opinion rather than an approved fact (Hyland, 2005). Contrary to previous studies that have reported the NNS researchers' relatively low use of hedging devices (Hyland, 2000b), the present study found somewhat diverging results. Hedges in the NS corpus in the present research consisted of 9 types and 135 tokens, whereas those of the NNS had 22 types and 201 tokens. Likewise, the NNS exceeded the NS in the use of hedges, and the patterns they used were somewhat dissimilar. Listed in the following table are hedges employed by the NS and the NNS.

Table 20. Hedges adopted by the NS and the NNS

	NS	NNS
Hedges	appears to be a can be interpreted as can be used to does not seem to it is not clear it is possible that it may be that may or may not seems to be a	can also be used can be accounted for can be interpreted as can be explained by can be separated by can be understood as can be used to can be used with if we assume that is assumed to be is expected to be is more likely to it is not clear it is possible that it seems that the may be due to reasonable to assume that the possibility that the this suggests that the under the assumption that we assume that the we predict that the

Surprisingly, the NNS not only adopted an extensive number of hedging devices but demonstrated a substantially varied use of the markers. The NNS took advantage of a wide variety of hedges from the high end of certainty to the lower ends (Biber, 2006). Such diversity was observed by various uses of verbs including *suggest*, *predict*, *expect*, and *assume* and adjectives such as *likely* and *possible*. On the other hand, lexical bundles including *may* were more often adopted by NS than NNS; in the NS corpus, 28 occurrences were found while there were 7 instances in the NNS corpus.

Upon an investigation of the concordances, the following structures mitigated the claims or inferences made by the writer. They were to protect the writer from any categorical assumptions and to open a space for readers to argue their interpretations.

(43) In English and many other languages, contrastive multiple focus can be expressed using normal word order with marked intonation, as in 30a. This *does not seem to* be the case for Indonesian; a cleft is normally used in these contexts, as illustrated in 30b. (NS026)

(44) *It may be that* there is some dialectal difference between Henry's informants and the Buckie and Portavogie speakers, but clearly the analysis developed above which ties nominative case on a pronoun to full agreement will not immediately extend. (NS042)

(45) This gradual rise can be interpreted as the result of an interpolation between the word-initial L tone and the accent H tone of the word-final syllable, with the penultimate syllable unspecified for tone. *This suggests that the* surface representation in NKK is sparsely, not fully, specified for tone. (NNS002)

(46) In child-directed speech, unmarked, nominative and accusative forms take 75%, 20%, and 5% of the occurrence, respectively (Albright 2008 citing I. Lee 1999). Thus, the same word *is more likely to* be produced in both the unmarked and accusative forms than in both the nominative and accusative forms. (NNS003)

In presenting results, the author in (43) expresses caution by employing hedges. Similarly, by using a mitigator, the author in (45) leaves room for readers to dispute in providing the author's analysis. In this way, authors can avoid possible criticisms that are related to their subjective judgments. In unfolding analysis, the author in (44) indicates another possible factor to the results. In (46), the author adopts the concept of likelihood in interpreting results. Such strategies show respect for the opinions of readers by signaling subjectivity and tentativeness of the authors' own views.

Self-mention is defined as the use of the first person pronouns and possessive forms so as to confine the author's claims, increase reliability, and develop credibility. It can build a reliable, intelligent and engaging colleague in the presentation of an authorial identity that accords with disciplinary conventions and the reflection of an effective degree of confidence and authority (Hyland, 2001a). The NS used 77 tokens of self-mention out of 5 types of lexical bundles, most of which presented the author's claims or provided a brief overview of the contents in introduction or the introductory part of a section (49, 50). On the other hand, the NNS employed 11 types of the bundles which occurred 131 times. Among them, listed in the following are *I*

have argued that (51), *I have shown that* (52), *I argue that the* (53), and *I propose that the* (54). While the former two from the NNS corpus provided a short summary of what has been discussed by the author with 24 times, the latter was used to introduce the author's arguments with 23 occurrences.

(47) And then in §3 *I will argue that* quantifying over individual concepts helps with the semantics of indefinite descriptions. The final section contains some concluding remarks. (NS025)

(48) *I will argue that* understanding the meaning of *It is clear that* p requires making a semantic distinction between main effect and side effects. This terminology comes from the theory of programming languages, but has fruitful application in natural languages. (NS028)

(49) In this section, *I have argued that* ta indicates that the antecedent is representing another representation. Additional meanings are pragmatically inferred. (NNS22)

(50) While SOR in Korean/Japanese seems to flout known constraints on A-movement, in particular, SSC/Relativized Minimality, *I have shown that* a closer investigation of the relevant facts allows us to maintain SSC as a constraint on A-movement. (NNS025)

(51) Building on this line of analysis, *I argue that the* (in)directness meaning in Korean evidential sentences is expressed by means of temporal relations between the evidence acquisition eventuality and the described eventuality. (NNS029)

(52) Given these cross-linguistic differences, *I propose that the* CMH can be restated as a parameter in UG for placing a reason adverb in the syntactic tree, as described in (62). (NNS035)

Although both the NS and the NNS commonly used self-mention strategies in presenting the author's claims, with the NNS using far more often, the two groups differed in the way to establish one's own arguments. Whereas the NS preferred to approach in a more indirect way such as anticipatory-*it* clauses, the NNS took advantage of the use of self-mention.

4.3.3.2 Engagement Bundles

Thompson and Thetela (1995) introduced the concept of engagement in "writer-in-the-text", as they regarded the participant to be responsible for the text. Although they look similar to other structures in their overuse by the NNS, there are some notable things to look at. As engagement bundles, those with a function of structuring appeared as 3 and 14 types in the NS and the NNS corpora, respectively. This aligns with the finding of this research that NNS are more prone to using metadiscursive bundles. In addition, the NS had a wider variety of usages of the inclusive first person pronoun *we*. The authors of the corpora in the present study adopted two types of engagement that were identified by Hyland (2001a) – the inclusive first person pronoun *we* and directives. Directives were employed in two types by each group, which are of too small a number to discuss. As for an inclusive first person pronoun *we*, the NS employed *as we will see* (53), *as we have seen* (54), *we have the following* (55), *we have already seen* (56), whereas the NNS adopted an inclusive first person pronoun *we* in *we have seen that* (57), *as we will see* (58) and *let us now consider*. Listed in the following are some of the examples from an inclusive first person *we*.

(53) *As we will see*, there are languages generated by one theory that are not generated by another;（NS008)

(54) *As we have seen*, however, this is exactly the situation that obtains in approaches that adopt the structural assumptions of weak layering. (NS006)

(55) It is as irrelevant to their syntax as, say, [past] is. Given this, *we have the following* set of pronouns in English: (NS046)

(56) Most of the examples of bukan V presented thus far involve negation of main clauses, but as *we have already seen* in examples 5 and 26, bukan V can also be used to negate some finite complement clauses. (NS052)

(57) *We have seen that* the NF approach may explain special status of nouns in some phonological domain. This raises a question as to whether the noun-verb asymmetry in Korean can also be explained by the NF approach. (NNS037)

(58) However, *as we will see,* not all tone languages preserve the input language prominence. (NNS008)

The concordances from the NNS corpus show that an inclusive first person pronoun is used when directing the reader from a brief summary of what has been presented thus far to the introduction of the writer's next strand of logic (57). They are also pointers to what is ahead (58). The native use of the inclusive *we* looks more diverse than this. Other than the overlapping bundle *we have seen that* which serves the same role, they guide the reader through the text by providing the reader a short overview of what follows next (53) and summarize what has been demonstrated in the study (54). They also introduce the upcoming principle, representation, properties, entities, definition, algorithm, derivation and lexicon (55), and enable the writer to unfold the discussion by specifically pointing to the data examined (56).

Directive is another device that appears in both corpora.

(59) In order to draw the connection between inventory shape and phonological activity which is one of the major motivations behind the theory of the contrastive hierarchy, *it is necessary to* assume that constraints motivating phonological processes operate over contrastively specified representations. (NS020)

(60) If I start calling that board in my garage a "structurally challenged table," pretty soon I may start to think of it as a table. But in this case *it is important to* bear

in mind the lesson (sometimes attributed to Abraham Lincoln) concerning how many legs a horse has if we call a tail a leg. (NS025)

(61) This raises the question of why there was no significant difference in the processing time for the clarification clause, whereas there was such a difference in the sensicality rating. Then *it should be noted that* the eye-tracking experiments were made on test items selected based on sensicality ratings and that their mean sensicality ratings showed no significant differences. (NNS023)

As shown above, the NS used directive markers to lead the reader into the selected line of reasoning by pointing the reader's attention to one particular direction (59, 60). The NNS manifested the use of a directive so as to steer the reader into the writer's arguments by emphasizing a main point (61).

4.3.3.3 Summary

As it was with other bundles, the NNS heavily overused participant-oriented bundles compared to the NS. According to previous research, NS adopted a variety of epistemic markers (Hyland & Milton, 1997) while NNS did not. Previous researchers also observed that NNS tended to express stronger commitments to assertions (Hyland & Milton, 1997). Surprisingly, the NNS in this study took advantage of a wide variety of hedging devices from the high end of certainty to the lower ends (Biber, 2006). In this study, NNS also used a variety of epistemic markers, but there was prevalent use of modal and lexical verbs. Previous research suggests that this is due to lower proficiency in NNS writers (Hyland & Milton, 1997). On the other hand, a substantial use of self-mention by NNS contrast with the findings of Hyland (2008a) in which Hong Kong writers were more reluctant to take responsibility for their ideas, as revealed through their underuse of personal pronouns. Possible explanations are that the control of the disciplinary nature as theoretical linguistics and the same degree of expertise between both groups in this study has been influential. Alternatively, the NS may have been more familiar and felt more comfortable with using predicative adjectives to convey the writer's own evaluation. Bundles

that indicated engagement also appeared far more frequently in the NNS corpus with less diversity.

5 Conclusion

5.1 Summary of Findings

This study attempted to (a) examine and compare the lexical bundles used by native and non-native English speakers, (b) investigate structural characteristics and differences in lexical bundles used between the two groups and (c) explore the different functional features of lexical bundles used between the groups.

It was found that the NNS used 2.7 times more bundles than the NS. A look at the 20 most common bundles by both groups suggests that the NNS are familiar with metadiscursive bundles. Structurally, both groups showed the similar tendency to favor PP-based, VP-based and NP-based bundles in order. Whereas the NNS displayed a substantial use of bundles, the NS showed variation in tense, subject, and modifier use. The NS also located some of the bundles to the places in which the NNS did not. The NNS preferred passive forms to active voice.

The NS and the NNS showed a similar overall trend in function; text-oriented bundles were the highest, research-oriented structures were next in common, and participant-oriented bundles were the least adopted. The NS relied more on research-oriented grouping bundles, which was not as popular among the NNS. It is not easy to generalize due to the relatively small number of the retrieved bundles, but we can assume that NS are more likely to use the patterns that refer to grouped individuals.

Both groups devoted a substantial proportion to text-oriented bundles, which is a distinctive feature of writers with high expertise. The NS used indirect devices to convey personal views, whereas the NNS preferred to use personal pronoun *I/we*. The NNS' reliance on metadiscursive bundles was notable in their frequent use of structuring bundles. They also presented results by comparing and contrasting with the literature. The experimental nature of articles in the NNS corpus has caused the higher number of causative bundles.

The NNS also exceedingly used participant-oriented bundles than the NS. Concerning stance markers, the NNS notably exhibited a varied use of hedges contrary to the literature.

While such diversity indicates their expertise, their prevailing use of modal and "speech act" verbs points to their non-native background. On the other hand, NNS enjoyed using self-mention devices, whereas the NS did not. Given that the phenomenon was negatively correlated with the use of the anticipatory-*it* patterns, it may be that predicative adjectives are not a possible option for the NNS. As for engagement bundles, the NS had a wider variety of use although they used them less often than the NNS.

The present thesis corroborates the previous studies that revealed the differences in use of lexical bundles between NS and NNS. On the other hand, the newly found similarities between the NS and the NNS researchers indicate that expertise is also an important element as much as nativeness.

5.2 Limitations and Future Studies

Although attempts were made to balance the number of texts between the NS and the NNS data, the final word counts were not balanced. Due to the availability of articles, the data pool of Korean English users was not as sufficient as the English native users'. There were also inconsistencies in the number of journals from which the articles were selected between the two groups, which may have influenced the results. Most importantly, a more comprehensive corpus could have validated the study for generalization.

A possible direction to future studies could be in applying the cross-disciplinary approach to subfields within linguistics. By expanding the scope from theoretical linguistics to theoretical and applied linguistics, researchers may be able to examine how the pattern and usage of lexical bundles between the two subfields differ within linguistics. It could reveal some of the different tendencies between theoretical and applied linguists and become a useful resource of wordlists in linguistics for EAP practitioners.

References

Ädel, A. (2006). *Metadiscourse in L1 and L2 English.* Amsterdam: John Benjamin Publishing.

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.

Anthony, L. (2014). AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/

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

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, *14*(3), 275-311.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

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

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

Chung, S., & Song, K. (2012). A corpus-based analysis of lexical bundles in academic articles on English literature, linguistics, and education written by English native speakers and Koreans. *Korean Journal of English Language and Linguistics*, *12*(4), 947-975.

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

Cowie, A.P. (Ed.). (1998). *Phraseology: Theory, Analysis and Applications*. Oxford: Oxford University Press.

Facchinetti, R. (Ed.). (2007). *Corpus Linguistics 25 Years On*. Amsterdam: Rodopi.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications* (pp. 145-160). Oxford: Clarendon.

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. (2000a). *Disciplinary Discourses: Social Interactions in Academic Writing.* London: Longman.

Hyland, K. (2000b). Hedges, boosters and lexical invisibility: noticing modifiers in academic texts. *Language Awareness*, *9*(4), 179-197.

Hyland, K. (2001a). Bringing in the reader: Addressee features in academic articles. *Written Communication*, *18*(4), 549-574.

Hyland, K. (2002b). Directives: Arguments and engagement in academic writing. *Applied Linguistics*, *23*(2), 215-239.

Hyland, K. (2005) Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, *7*(2), 173-192.

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 variation. *English for Specific Purposes*, *27*(1), 4-21.

Hyland, K. & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing*. *16*(2), 183-205.

Lee, S. (2012). A study on the use of lexical bundles in science journals. In *Proceedings of the 40th Biannual Conference of the Linguistic Association of Korea*, Seoul, Korea.

Lindquist, H. (2009). *Corpus Linguistics and the Description of English*. Edinburgh: Edinburgh University Press.

Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In Sinclair, J. McH. (Ed.), (2004). *How to Use Corpora in Language Teaching*. (pp.125-152). Amsterdam: John Benjamins Publishing.

Pawley, A. & Syder, F. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In Richards, J. C. & Schmidt, R. W. (Eds.), *Language and Communication* (pp.191-226). New York: Longman.

Richards, J. C. & Schmidt, R. W. (Eds.). (2010). *Language and Communication*. New York: Longman.

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

Sag, I. A., Baldwin, T., Bond, F., Copestake, A., & Flickinger, D. (2002). Multiword expressions: A pain in the neck for NLP. In Proceedings of the third international conference on computational linguistics and intelligent text processing (CICLing-2002), *Lecture Notes in Computer Science*, London, UK, Vol. 2276 (pp. 1–15).

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.

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

Sinclair, J. (Ed.). (2004). *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins Publishing.

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

Swales, J. M. (1985), English language papers and authors' first language: preliminary explorations, *Scientometrics*, *8*(1-2), 91-101.

Thompson, G., & Thetela, P. (1995). The sound of one hand clapping: The management of interaction in academic discourse. *Text, 15*(1), 103-127.

Vassileva, I. (1998). "Who am I / who are we in academic writing? A contrastive analysis of authorial presence in English, German, French, Russian and Bulgarian". *International Journal of Applied Linguistics, 8*(2), 163-190.

Appendices

A. Complete list of NS bundles

Rank	Freq.	Range	Lexical Bundle	Structure	Function 1	Function 2
1	69	27	on the other hand	PP	Comparative	
2	60	21	in the case of	PP+*of*	Framing	
3	57	24	on the basis of	PP+*of*	Framing	
4	50	22	with respect to the	PP	Framing	
5	46	19	to account for the	V/A+*to*	Objective	
6	39	25	the fact that the	V/N+*that*	Framing	
7	39	15	the rest of the	NP+*of*	Grouping	
8	37	8	the left edge of	NP+*of*	Location	
9	35	20	in addition to the	PP	Additive	
10	33	22	as well as the	Others	Additive	
11	33	17	in terms of the	PP	Framing	
12	33	15	at the level of	PP	Description	
13	33	12	it is clear that	Anticipatory *it*	Stance	Inferential
14	32	21	in this case the	PP	Framing	
15	32	20	it is possible to	Anticipatory *it*	Inferential	
16	32	20	that there is a	VP+*that*	Description	
17	30	18	in a way that	PP	Framing	
18	30	14	at the same time	PP	Additive	
19	29	17	in the sense that	PP	Framing	
20	28	16	at the end of	PP+*of*	Location	
21	28	13	in the analysis of	PP+*of*	Procedure	
22	25	15	a subset of the	NP+*of*	Grouping	
23	24	15	the nature of the	NP+*of*	Description	
24	24	13	that it is not	VP+*that*	Inferential	
25	24	12	in the sense of	PP+*of*	Framing	
26	23	12	in the context of	PP+*of*	Framing	
27	23	8	the set of all	NP+*of*	Grouping	
28	22	19	in the same way	PP	Comparative	
29	22	17	the structure of the	NP+*of*	Description	
30	22	16	by the fact that	PP	Framing	
31	22	16	it is important to	Anticipatory *it*	Stance	Engagement
32	22	6	at the left edge	PP	Location	
33	21	15	as we have seen	*I/We*+V	Engagement	
34	21	14	in other words the	PP	Framing	

35	21	11	in the previous section	PP	Structuring	
36	21	7	of the examples in	PP+*of*	Structuring	
37	20	17	that there is no	V/N+*that*	Procedure	
38	20	16	in the absence of	PP+*of*	Framing	
39	20	14	it is necessary to	Anticipatory *it*	Engagement	Inferential
40	20	14	it is not clear	Anticipatory *it*	Stance	Inferential
41	20	12	in the next section	PP	Structuring	
42	20	5	the effects of the	NP+*of*	Causative	
43	20	4	is a matter of	Copula *be*	Description	
44	19	13	the full range of	NP+*of*	Grouping	
45	19	11	the presence of a	NP+*of*	Description	
46	19	11	the use of the	NP+*of*	Procedure	
47	18	10	for the sake of	PP+*of*	Causative	
48	18	10	the role of the	NP+*of*	Description	
49	17	12	to the extent that	PP	Description	
50	17	11	it is possible that	Anticipatory *it*	Stance	Inferential
51	17	10	if and only if	Adv. cl	Framing	
52	17	10	in this section i	PP	Structuring	
53	17	10	the meaning of the	NP+*of*	Description	
54	16	15	can be used to	V/A+*to*	Inferential	Stance
55	16	13	in terms of a	PP+*of*	Framing	
56	16	12	from the fact that	PP	Framing	
57	16	12	the question of whether	NP+*of*	Framing	
58	16	11	on the one hand	PP	Comparative	
59	16	10	may or may not	Other VP	Stance	Inferential
60	16	10	to the left of	PP+*of*	Location	
61	16	9	in this section we	PP	Structuring	
62	16	8	the content of the	NP+*of*	Procedure	
63	16	5	at the right edge	PP	Location	
64	16	5	the degree to which	NP	Description	
65	15	13	turns out to be	V/A+*to*	Inferential	
66	15	12	in the form of	PP+*of*	Framing	
67	15	11	to be able to	V/A+*to*	Description	
68	15	10	as we will see	*I/We*+V	Engagement	
69	15	10	i will argue that	*I/We*+V	Stance	Inferential
70	15	9	a special case of	NP+*of*	Framing	
71	15	9	can be seen in	Passive	Structuring	Engagement
72	15	9	the beginning of the	NP+*of*	Procedure	
73	15	9	to the set of	PP+*of*	Grouping	
74	15	4	the probability of the	NP+*of*	Quantification	

75	14	12	is based on the	Passive	Framing	
76	14	11	seems to be a	V/A+*to*	Stance	Inferential
77	14	9	it is not a	Anticipatory *it*	Inferential	
78	14	9	the same way as	NP	Comparative	
79	14	8	can be interpreted as	Passive	Inferential	Stance
80	14	8	on the assumption that	PP	Framing	
81	14	8	to the right of	PP+*of*	Location	
82	14	8	we have already seen	*I/We*+V	Engagement	Structuring
83	14	7	it is easy to	Anticipatory *it*	Stance	Inferential
84	14	6	that there is some	V/A+*that*	Procedure	
85	14	6	the source of the	NP+*of*	Causative	
86	14	5	the right edge of	NP+*of*	Location	
87	14	4	of the theory of	PP+*of*	Framing	
88	13	11	appears to be a	V/A+*to*	Stance	Inferential
89	13	10	does not seem to	V/A+*to*	Stance	Inferential
90	13	8	such as those in	PP	Structuring	
91	13	7	along the lines of	PP+*of*	Comparative	
92	12	12	to be the case	V/A+*to*	Framing	
93	12	11	in a variety of	PP+*of*	Grouping	
94	12	10	the same as the	Adj. ph	Comparative	
95	12	9	as far as i	Adv. cl	Framing	
96	12	9	as part of the	PP+*of*	Grouping	
97	12	9	it may be that	Anticipatory *it*	Stance	Inferential
98	12	9	the way in which	NP	Framing	
99	12	9	this is illustrated in	Passive	Structuring	
100	12	9	this is not the	Copula *be*	Inferential	
101	12	8	the interpretation of the	NP+*of*	Inferential	
102	12	8	the position of the	NP+*of*	Description	
103	12	8	this is the case	Copula *be*	Framing	
104	12	7	we have the following	*I/We*+V	Engagement	Structuring
105	12	6	of the same type	PP+*of*	Grouping	
106	12	4	be thought of as	Passive	Inferential	
107	12	3	to the right edge	PP	Location	

B. Complete list of NNS bundles

Rank	Freq.	Range	Lexical Bundle	Structure	Function 1	Function 2
1	147	33	on the other hand	PP	Comparative	
2	48	9	as a function of	PP+*of*	Framing	
3	45	16	with respect to the	PP	Framing	
4	45	12	in the present study	PP	Structuring	
5	41	17	in terms of the	PP+*of*	Framing	
6	34	12	in this section i	PP	Structuring	
7	32	20	in the case of	PP+*of*	Framing	
8	28	15	the results of the	NP+*of*	Causative	
9	27	15	in the next section	PP	Structuring	
10	27	5	the duration of the	NP+*of*	Quantification	
11	26	17	at the same time	PP	Framing	
12	26	7	the use of the	NP+*of*	Description	
13	25	25	the fact that the	V/N+*that*	Framing	
14	23	15	as well as the	Others	Additive	
15	23	15	in other words the	PP	Framing	
16	21	10	in the previous section	PP	Structuring	
17	21	9	we have seen that	*I/We*+V	Stance	Inferential
18	20	10	on the basis of	PP+*of*	Framing	
19	20	9	prediction is borne out	Passive	Inferential	
20	20	3	we expect that the	*I/We*+V	Stance	Inferential
21	19	12	as shown in the	Adv. cl	Structuring	Engagement
22	19	11	as a result of	PP+*of*	Causative	
23	19	9	as can be seen	Adv. cl	Engagement	
24	19	8	significant main effect of	NP+*of*	Causative	
25	19	6	the location of the	NP+*of*	Location	
26	18	11	at the end of	PP+*of*	Location	
27	18	10	of the present study	PP+*of*	Structuring	
28	18	9	i have argued that	*I/We*+V	Stance	
29	17	3	cannot be separated by	Passive	Inferential	Stance
30	16	14	from the fact that	PP	Framing	
31	16	12	due to the fact	PP	Causative	
32	16	12	that there is a	V/N+*that*	Inferential	
33	16	11	can be interpreted as	Passive	Inferential	Stance
34	16	9	not the case that	V/N+*that*	Framing	
35	16	7	i argue that the	*I/We*+V	Inferential	Stance
36	16	3	a change of state	NP+*of*	Description	
37	15	12	the relationship between	NP	Comparative	

			the			
38	15	9	difference between the two	NP	Comparative	
39	15	9	in line with the	PP	Comparative	
40	15	7	in this paper i	PP	Structuring	
41	15	3	a state of affairs	NP+*of*	Description	
42	15	3	condition than in the	NP	Comparative	
43	14	9	by the fact that	PP	Causative	
44	14	9	in this section we	PP	Structuring	
45	14	7	used to refer to	V/A+*to*	Description	
46	14	6	as shown in fig	Adv. cl	Structuring	Engagement
47	13	9	the difference between the	NP	Comparative	
48	13	8	despite the fact that	PP	Framing	
49	13	8	in the sense of	PP+*of*	Framing	
50	13	8	the size of the	NP+*of*	Quantification	
51	13	7	is more likely to	V/A+*to*	Stance	Inferential
52	13	7	this suggests that the	V/N+*that*	Inferential	Stance
53	13	6	there was a significant	Copula *be*	Inferential	
54	13	4	as shown in figure	Adv. cl	Structuring	Engagement
55	13	3	we predict that the	*I/We*+V	Inferential	Stance
56	12	10	in addition to the	PP	Additive	
57	12	9	a significant main effect	NP	Causative	
58	12	9	in contrast to the	PP	Comparative	
59	12	9	in such a way	PP	Framing	
60	12	9	is responsible for the	Copula *be*	Causative	
61	12	7	first and the second	Adj. ph	Location	
62	12	5	can be accounted for	Passive	Causative	Stance
63	12	4	in the examples in	PP	Structuring	
64	12	3	can be separated by	Passive	Procedure	Stance
65	11	9	will be discussed in	Passive	Structuring	
66	11	8	as shown in table	Adv. cl	Structuring	Engagement
67	11	8	does not have a	Active	Description	
68	11	8	if we assume that	Adv. cl	Engagement	Stance
69	11	8	it should be noted	Passive	Stance	Engagement
70	11	7	the present study has	Active	Structuring	
71	11	7	the two types of	NP+*of*	Grouping	
72	11	6	the present study however	NP	Structuring	
73	11	6	this is illustrated in	Passive	Structuring	
74	11	4	a unified account for	NP	Causative	
75	11	3	a situation in which	NP	Framing	
76	11	3	for the occurrence of	PP+*of*	Causative	

77	11	3	it will be shown	Anticipatory *it*	Structuring	
78	10	8	as far as the	Adv. cl	Framing	
79	10	8	that there is no	V/N+*that*	Inferential	
80	10	7	it is possible that	Anticipatory *it*	Stance	Inferential
81	10	7	the same way as	Adj. ph	Comparative	
82	10	6	the meaning of the	NP+*of*	Description	
83	10	5	is due to the	Copula *be*	Causative	
84	10	4	as in the example	Adv. cl	Structuring	Engagement
85	10	4	at the right edge	PP	Location	
86	10	4	in the framework of	PP+*of*	Framing	
87	10	4	is followed by a	Passive	Structuring	
88	10	3	such as that in	PP	Structuring	
89	9	9	for example in the	PP	Structuring	Engagement
90	9	8	as opposed to the	Adv. cl	Comparative	
91	9	8	the claim that the	V/N+*that*	Framing	
92	9	7	for the purpose of	PP+*of*	Procedure	
93	9	7	in the context of	PP+*of*	Framing	
94	9	7	to the extent that	PP	Description	
95	9	6	the use of a	NP+*of*	Description	
96	9	6	as one of the	PP+*of*	Grouping	
97	9	6	in this paper we	PP	Structuring	
98	9	6	is based on the	Copula *be*	Framing	
99	9	6	play a role in	Active	Causative	
100	9	6	that it is not	V/N+*that*	Inferential	
101	9	5	as was the case	Adv. cl	Comparative	Engagement
102	9	5	in order to examine	Adv. cl	Objective	
103	9	5	no main effect of	NP+*of*	Causative	
104	9	5	on the nature of	PP+*of*	Description	
105	9	5	that the degree of	V/N+*that*	Description	
106	9	5	under the assumption that	PP	Framing	Stance
107	9	4	can also be used	Passive	Inferential	Stance
108	9	4	can be used with	Passive	Description	Stance
109	9	4	in the use of	PP+*of*	Description	
110	9	4	the target of the	NP+*of*	Procedure	
111	9	4	was found to be	V/A+*to*	Inferential	
112	9	3	is used with the	Copula *be*	Citation	
113	9	3	of the effect of	PP+*of*	Causative	
114	9	3	on the type of	PP+*of*	Description	
115	9	3	the left edge of	NP+*of*	Location	
116	8	8	the present study is	Copula *be*	Structuring	

117	8	8	a result of the	NP+*of*	Causative	
118	8	8	as well as in	Others	Additive	
119	8	8	on the one hand	PP	Comparative	
120	8	8	similar to that of	Adj. ph	Comparative	
121	8	8	the assumption that the	V/N+*that*	Framing	
122	8	7	be attributed to the	Passive	Causative	Inferential
123	8	7	there was no significant	Copula *be*	Inferential	
124	8	6	at the beginning of	PP+*of*	Location	
125	8	6	at the time of	PP+*of*	Procedure	
126	8	6	one of the most	NP+*of*	Grouping	
127	8	5	are given in the	Passive	Structuring	
128	8	5	are summarized in table	Passive	Structuring	
129	8	5	it is predicted that	Anticipatory *it*	Inferential	
130	8	5	that the present study	V/N+*that*	Structuring	
131	8	5	the extent to which	NP	Description	
132	8	5	the possibility that the	V/N+*that*	Stance	Inferential
133	8	4	be thought of as	Passive	Inferential	
134	8	4	cannot be used in	Passive	Inferential	
135	8	4	contrary to the facts	PP	Comparative	
136	8	4	in a situation where	PP	Framing	
137	8	4	in the current study	PP	Structuring	
138	8	4	it is notable that	Anticipatory *it*	Stance	
139	8	3	as illustrated in the	Adv. cl	Structuring	Engagement
140	8	3	at the edges of	PP+*of*	Location	
141	8	3	the source of the	NP+*of*	Causative	
142	7	8	the nature of the	NP+*of*	Description	
143	7	7	be accounted for by	Passive	Causative	Inferential
144	7	7	for the sake of	PP+*of*	Causative	
145	7	7	in the following section	PP	Structuring	
146	7	7	is assumed to be	V/A+*to*	Stance	Inferential
147	7	7	is attributable to the	Copula *be*	Causative	Inferential
148	7	7	may be due to	Copula *be*	Stance	Causative
149	7	6	to that of the	PP+*of*	Comparative	
150	7	6	the right edge of	NP+*of*	Location	
151	7	6	as a result the	PP	Causative	
152	7	6	as to whether the	PP	Framing	
153	7	6	as we will see	*I/We*+V	Engagement	
154	7	6	different from that of	Adj. ph	Comparative	
155	7	6	in the vicinity of	PP+*of*	Location	
156	7	6	is expected to be	V/A+*to*	Stance	Inferential

157	7	6	it has been observed	Anticipatory *it*	Citation	
158	7	6	it is expected that	Anticipatory *it*	Inferential	
159	7	6	it seems that the	Anticipatory *it*	Stance	Inferential
160	7	6	presence or absence of	NP+*of*	Description	
161	7	6	the result of the	NP+*of*	Causative	
162	7	6	to account for the	V/A+*to*	Objective	
163	7	6	with the fact that	PP	Framing	
164	7	5	as far as i	Adv. cl	Framing	
165	7	5	can be explained by	Passive	Stance	Inferential
166	7	5	do not have a	Active	Description	
167	7	5	in favor of the	PP+*of*	Framing	
168	7	5	it is hard to	Anticipatory *it*	Stance	Inferential
169	7	5	it is possible to	Anticipatory *it*	Inferential	
170	7	5	the first and second	Adj. ph	Location	
171	7	5	the frequency of the	NP+*of*	Quantification	
172	7	5	the length of the	NP+*of*	Quantification	
173	7	5	the majority of cases	NP+*of*	Grouping	
174	7	5	the position of the	NP+*of*	Location	
175	7	5	this is because the	Copula *be*	Causative	
176	7	5	when there is no	Adv. cl	Framing	
177	7	4	an analysis of the	NP+*of*	Procedure	
178	7	4	an increase in the	NP	Quantification	
179	7	4	as described in the	Adv. cl	Structuring	Engagement
180	7	4	belong to the same	Active	Procedure	
181	7	4	can be used to	V/A+*to*	Inferential	Stance
182	7	4	discussed in the previous	Other VP	Structuring	
183	7	4	do not have to	Active	Citation	
184	7	4	further support for the	NP	Inferential	
185	7	4	i propose that the	*I/We*+V	Inferential	Stance
186	7	4	in the state of	PP+*of*	Description	
187	7	4	is a set of	Copula *be*	Grouping	
188	7	4	is attached to the	Passive	Citation	
189	7	4	is motivated by the	Passive	Inferential	
190	7	4	is the same as	Copula *be*	Comparative	
191	7	4	it is important to	Anticipatory *it*	Stance	Engagement
192	7	4	let us first consider	Other VP	Engagement	Structuring
193	7	4	present study is to	V/A+*to*	Structuring	
194	7	4	regardless of whether the	PP+*of*	Framing	
195	7	4	the influence of the	NP+*of*	Causative	
196	7	4	the presence of the	NP+*of*	Description	

197	7	4	the results showed that	V/N+*that*	Inferential	
198	7	4	to do with the	V/A+*to*	Framing	
199	7	3	all else being equal	NP	Procedure	
200	7	3	as reflected in the	Adv. cl	Engagement	Structuring
201	7	3	examples are provided in	Passive	Structuring	
202	7	3	in the interpretation of	PP+*of*	Framing	
203	7	3	the content of the	NP+*of*	Description	
204	7	3	the same type of	NP+*of*	Grouping	
205	7	3	was observed in the	Passive	Inferential	
206	6	7	the interpretation of the	NP+*of*	Inferential	
207	6	6	as pointed out by	Adv. cl	Citation	
208	6	6	goal of this paper	NP+*of*	Structuring	
209	6	6	has a three way	Active	Description	
210	6	6	in the majority of	PP+*of*	Grouping	
211	6	6	is supported by the	Passive	Causative	
212	6	6	it is not clear	Anticipatory *it*	Stance	
213	6	6	participated in the experiment	Active	Procedure	
214	6	6	the analysis of the	NP+*of*	Procedure	
215	6	6	the case with the	NP	Framing	
216	6	6	the present study therefore	NP	Structuring	
217	6	6	the rest of the	NP+*of*	Grouping	
218	6	6	this is not the	Copula *be*	Inferential	
219	6	5	as in the case	Adv. cl	Comparative	Engagement
220	6	5	at least in part	PP	Description	
221	6	5	but they do not	Other VP	Inferential	
222	6	5	cannot be attributed to	Passive	Causative	Inferential
223	6	5	depending on the context	PP	Inferential	
224	6	5	in such a case	PP	Framing	
225	6	5	it is difficult to	Anticipatory *it*	Stance	
226	6	5	shed light on the	Active	Inferential	
227	6	5	since there is no	Adv. cl	Inferential	
228	6	5	the case that the	V/N+*that*	Framing	
229	6	5	the view that the	V/N+*that*	Framing	
230	6	5	to be associated with	Copula *be*	Citation	
231	6	5	we assume that the	*I/We*+V	Inferential	Stance
232	6	5	we propose that the	*I/We*+V	Inferential	Stance
233	6	4	to examine whether the	V/A+*to*	Objective	
234	6	4	a crucial role in	NP	Causative	
235	6	4	about cm from the	Adj. ph	Quantification	

236	6	4	about the nature of	PP+*of*	Description	
237	6	4	and the degree of	NP+*of*	Description	
238	6	4	are shown in table	Passive	Structuring	
239	6	4	as compared to the	Adv. cl	Comparative	
240	6	4	as discussed in section	Adv. cl	Structuring	Engagement
241	6	4	does not have to	Other VP	Inferential	
242	6	4	goal of the present	NP+*of*	Structuring	
243	6	4	however there was no	Copula *be*	Inferential	
244	6	4	in other words it	PP	Citation	
245	6	4	is part of the	Copula *be*	Citation	
246	6	4	it has been shown	Anticipatory *it*	Citation	
247	6	4	it has been suggested	Anticipatory *it*	Description	
248	6	4	on any of the	PP+*of*	Grouping	
249	6	4	on the degree of	PP+*of*	Description	
250	6	4	reasonable to assume that	Adj. ph	Inferential	Stance
251	6	4	speakers were asked to	Passive	Procedure	
252	6	4	supporting evidence for the	NP	Inferential	
253	6	4	the difference in the	NP	Quantification	
254	6	4	the framework of a	NP+*of*	Framing	
255	6	4	the perspective of the	NP+*of*	Framing	
256	6	4	turned out to be	V/A+*to*	Inferential	
257	6	4	was significant only in	Copula *be*	Inferential	
258	6	3	be referred to as	Passive	Description	
259	6	3	be related to the	Passive	Inferential	
260	6	3	can be understood as	Passive	Stance	Citation
261	6	3	i have shown that	*I/We*+V	Stance	
262	6	3	let us now consider	Other VP	Structuring	Engagement
263	6	3	no support for the	NP	Inferential	
264	6	3	over the course of	PP+*of*	Procedure	
265	6	3	significant main effects of	NP+*of*	Causative	
266	6	3	the present study also	NP	Structuring	
267	6	3	there was also a	Copula *be*	Inferential	
268	6	3	were divided into two	Passive	Procedure	
269	6	3	whether and how the	Adv. cl	Objective	
270	6	3	with the exception of	PP+*of*	Framing	

C. Selected journals for the NS and the NNS subcorpora

Subdiscipline	NS	NNS
Phonology/Phonetics	Journal of Phonetics Language Linguistic Inquiry Phonology	Journal of Acoustical Society of America Journal of East Asian Linguistics Journal of Phonetics Journal of Voice Language and Cognitive Processes Lingua Phonetica Phonology
Syntax	English Language and Linguistics Journal of Linguistics Journal of Logic and Computation Language Information Language & Linguistics Compass Language, Language and Communication Lingua Linguistic Inquiry Linguistics and Philosophy Mind & Language Natural Language and Linguistic Theory	Journal of East Asian Linguistics Lingua Linguistic Inquiry Linguistic Theory Natural Language and Linguistic Theory
Semantics/Pragmatics	Inquiry Language Linguistics and Philosophy Linguistic and Philosophical Investigations Mind and Language Philosophical Perspectives Research on Language and Computation	Journal of Cognitive Science Journal of East Asian Linguistics Journal of Pragmatics Journal of Semantics Language Sciences Research on Language & Computation

D. Complete list of NS articles

Phonology/Phonetics

Albright, A. (2009). Feature-based generalization as a source of gradient acceptability. *Phonology*, *26*(1), 9-41.

Bennett, R. (2013). The uniqueness of metrical structure: rhythmic phonotactics in Huariapano. *Phonology*, *30*(3), 355-398.

Blust, R. (2012). One mark per word? Some patterns of dissimilation in Austronesian and Australian languages. *Phonology*, *29*(3), 355-381.

Buckley, E. (2009). Locality in metrical typology. *Phonology*, *26*(4), 389-435.

Carpenter, A. (2010). A naturalness bias in learning stress, *Phonology*, *27*(3), 345-392.

Flack, K. (2009). Constraints on onsets and codas of words and phrases. *Phonology*, *26*(2), 269-302.

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, *39*(3), 379-440.

Hayes, B., & White, J. (2013). Phonological naturalness and phonotactic learning. *Linguistic Inquiry*, *44*(1), 45-75.

Hyde, B. (2012). The odd-parity input problem. *Phonology*, *29*(3), 383-431.

Kaplan, A. (2011). Variation through markedness suppression. *Phonology*, *28*(3), 331-370.

Labov, W. (2006). A sociolinguistic perspective on sociophonetic research. *Journal of Phonetics*, *34*(4), 500-514.

Mackenzie, S. (2013). Laryngeal co-occurrence restrictions in Aymara: Contrastive representations and constraint interaction. *Phonology*, *30*(2), 297-345.

Pater, J. (2008). Gradual learning and convergence. *Linguistic Inquiry*, *39*(2), 334-345.

Pater, J. (2014). Canadian raising with language-specific weighted constraints. *Language*, *90*(1), 230-240.

Pycha, A. (2010). A test case for the phonetics-phonology interface: Gemination restrictions in Hungarian. *Phonology*, *27*(1), 119-152.

Rose, S., & Walker, R. (2004). A typology of consonant agreement as correspondence. *Language*, *80*(3), 475-531.

Ryan, K. (2011). Gradient syllable weight and weight universals in quantitative metrics. *Phonology*, *28*(3), 413-454.

Ryan, K. (2014). Onsets contribute to syllable weight: statistical evidence from stress and meter. *Language*, *90*(2), 309-341.

Tucker, B. (2011). The effect of reduction on the processing of flaps and /g/ in isolated words, *Journal of Phonetics*, *39*(3), 312-318.

Syntax

Adger, D. (2006). Combinatorial variability**.** *Journal of Linguistics, 42*(3), 503-530.

Adger, D., & Trousdale, G. (2007). Variation in English syntax: Theoretical implications. *English Language and Linguistics*, *11*(2), 261-278.

Adger, D., & Smith J. (2010). Variation in agreement: A lexical feature-based approach. *Lingua, 120*(5), 1109-1134.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry, 36*(1), 1-22.

Chomsky, N. (2013). Problems of projection. *Lingua, 130*, 33-49.

Goldberg, A., & Jackendoff, R. (2004). The English resultative as a family of constructions. *Language*, *80*(3), 532-568.

Harbour, D. (2011). Mythomania? Methods and morals from 'The myth of language universals'. *Lingua, 121*(12), 1820-1830.

Kroeger, P. (2014). External negation in Malay/Indonesian. *Language, 90*(1), 137-184.

Pollard, C. (2007). Hyperintensions. *Journal of Logic and Computations, 18*(2), 257-281.

Pullum, G. (2011). On the mathematics of syntactic structures. *Journal of Logic, Language and Information, 20*(3), 277-296.

Pullum, G. (2013a). Consigning phenomena to performance: A response to Neeleman. *Mind and Language, 28*(4), 532-537.

Pullum, G. (2013b). The central question in comparative syntactic metatheory. *Mind and Language, 28*(4), 492-521.

Pullum, G. (2014). Fear and loathing of the English passive. *Language and Communication, 37*, 60-74.

Pullum, G., & Kyle, R. (2007). Argument or no argument? *Linguistics and Philosophy, 30*(2), 277-287.

Sabbagh, J. (2007). Ordering and linearizing rightward movement. *Natural Language & Linguistic Theory, 25*(2), 349-401.

Sabbagh, J. (2011). Adjectival passives and the structure of VP in Tagalog. *Lingua, 121*(8), 1424-1452.

Sabbagh, J. (2014). Right node raising. *Language and Linguistics Compass, 8*(1), 24-35.

Semantics/Pragmatics

Abbott, B. (2011). Support for individual concepts. *Linguistic and Philosophical Investigations, 10*, 23-44.

Anderson, J. (2004). On the grammatical status of names. *Language, 80*(3), 435-474.

Barker, C. (2005). Remark on Jacobson 1999: Crossover as a local constraint. *Linguistics and Philosophy, 28*(4), 447–472.

Barker, C. (2007). Parasitic scope. *Linguistics and Philosophy, 30*(4), 407-444.

Barker, C. (2009). Clarity and the grammar of skepticism. *Mind and Language, 24*(3), 253-273.

Barker, C. (2013). Scopability and sluicing. *Linguistics and Philosophy*, *36*(3)**,** 187-223.

Braun, D. (2012). An invariantist theory of 'might' might be right. *Linguistics and Philosophy*, *35*(6), 461-489.

Burnett, H. (2014). A delineation solution to the puzzles of absolute adjectives. *Linguistics and Philosophy*, *37*(1), 1-39.

Charlow, N. (2013). Conditional preferences and practical conditionals. *Linguistics and Philosophy*, *36*(6), 463-511.

Copestake, A., Flickinger, D., Pollard, C., & Sag, I. (2005). Minimal recursion semantics: An introduction. *Research on Language and Computation*, *3*(4), 281-332.

Glanzberg, M. (2008). Quantification and contributing objects to thoughts. *Philosophical Perspectives*, *22*(1), 207-231.

Kertz, L. (2013). Verb phrase ellipsis: The view from information structure. *Language*, *89*(3), 390-428.

Lasersohn, P. (2012). Contextualism and compositionality. *Linguistics and Philosophy*, *35*(2), 171-189.

Lasersohn, P. (2013). Non-world indices and assessment-sensitivity. *Inquiry*, 56(2-3), 122-148.

Steedman, M. (2014). The surface-compositional semantics of English intonation: Audio files accompanying examples. *Language*, *90*(1), 2-57.

Whitman, N. (2004). Semantics and pragmatics of English verbal dependent coordination. *Language*, *80*(3), 403-434.

E. Complete list of NNS articles

Phonology/Phonetics

Cho, H. (2011). The timing of phrase-initial tones in Seoul Korean: a weighted-constraint model. *Phonology, 28*(3), 293-330.

Cho, T. (2004). Prosodically-conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics, 32*(2), 141-176.

Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a,i/ in English. *Journal of the Acoustical Society of America, 117*(6), 3867-3878.

Cho, T. (2008). Prosodic strengthening in transboundary V-to-V lingual movement in American English. *Phonetica, 65* (1-2), 45-61.

Cho, T., Lee, Y., & Kim, S. (2011). Communicatively driven versus prosodically driven hyper-articulation in Korean. *Journal of Phonetics, 39*(3), 344-361.

Cho, T., Kim, J., & Kim, S. (2013). Preboundary lengthening and preaccentual shortening across syllables in a trisyllabic word in English. *Journal of the Acoustical Society of America, 133*(5), 384-390.

Cho, T., Yoon, Y., & Kim, S. (2014). Effects of prosodic boundary and syllable structure on the temporal realization of CV gestures in Korean. *Journal of Phonetics, 44*, 96-109.

Kim, S., & Cho, T. (2013). Prosodic boundary information modulates phonetic categorization. *Journal of the Acoustical Society of America, 134*(1), 19-25.

Jun, J. (2010). Stem-final obstruent variation in Korean. *Journal of East Asian Linguistics. 19*(2), 137-179.

Jun, J., Kim, J., Lee, H., & Jun, S. (2006). The Prosodic structure and pitch accent of Northern Kyungsang Korean. *Journal of East Asian Linguistics*, *15*(4), 289-317.

Jun, Sun-Ah (2010) The implicit prosody hypothesis and overt prosody in English. *Language and Cognitive Processes, 25*(7), 1201-1233.

Kang, Y. (2010). The emergence of phonological adaptation from phonetic adaptation: English loanwords in Korean. *Phonology, 27*(2), 225-253.

Kang, Y. (2010). Tutorial overview: Suprasegmental adaptation in loanwords. *Lingua, 120*(9), 2295-2310.

Kang, Y. (2013). L1 phonotactic restrictions and perceptual adaptation: English affricates in Contemporary Korean. *Journal of East Asian Linguistics, 22*(1), 39-63.

Kang, Y., & Han, S. (2013). Tonogenesis in early Contemporary Seoul Korean: A longitudinal case study. *Lingua, 134*, 62-74.

Moon, K., Chung, S., Park, H., & Kim, H. (2012). Materials of acoustic analysis: Sustained vowel versus sentence. *Journal of Voice, 26*(5), 563-565.

Son, M., Kim, S., & Cho, J. (2012). Supralaryngeal articulatory signatures of three-way

contrastive labial stops in Korean. *Journal of Phonetics, 40*(1), 92-108.

Syntax

Chung, D. (2005). What does bare-*ko* coordination tell about post-verbal morphology in Korean? *Lingua, 115*(4), 549-568.

Han, C., & Kim, J. (2004). Are there double relative clauses in Korean? *Linguistic Inquiry, 35*(2), 315-337.

Han, C., & Lee, C. (2007). On negative imperatives in Korean. *Linguistic Inquiry, 38*(2), 373-395.

Ko, H. (2005). Syntax of why-in-situ: Merge into [Spec,CP] in the overt syntax. *Natural Language & Linguistic Theory, 23*(4), 867-916.

Ko, H. (2006). Base-output correspondence in Korean nominal inflection. *Journal of East Asian Linguistics 15*(3), 195-243.

Ko, H. (2007). Asymmetries in scrambling and cyclic linearization. *Linguistic Inquiry, 38*(1), 49-83.

Ko, H. (2011). Predication and edge effects. *Natural Language and Linguistic Theory, 29*(3), 725-778.

Yoon, S. (2011). A structural asymmetry in intervention effects. *Lingua, 121*(5), 942-962.

Yoon, S. (2013). Parametric variation in subordinate expletive negation: Japanese/Korean vs. others. *Journal of East Asian Linguistics, 22*(2), 133-166.

Semantics/Pragmatics

Kwon, I. (2012). Please confirm what I inferred: On the Korean evidential marker -Napo-. *Journal of Pragmatics, 44*(7), 958-969.

Kwon, I. (2014). The viewpoint shifter -eha- in Korean and how it affects causal event structure. *Language Sciences, 42*(2), 30-42.

Lee, E. (2010). Pluperfects in Korean and English discourse. *Journal of Pragmatics, 42*(3), 766-780.

Lee, E., & Kim. H. (2007). On cross-linguistic variations in imperfective aspect: The case of L2 Korean. *Language Learning, 57*(4), 651-685.

Lee, E., & Choi, J. (2009). Two nows in Korean. *Journal of Semantics, 26*(1), 87-107

Lee, J. (2013). Temporal constraints on the meaning of evidentiality. *Natural Language Semantics, 21*(1), 1-41.

Lee, H. (2005). Presupposition and implicature under negation. *Journal of Pragmatics, 37*, 595-609.

Lee, H. (2010). Referring expressions in English and Korean political news. *Journal of Pragmatics, 42*(9), 2506-2518.

Lee, H. (2011). The use of the double form in Korean discourse. *Journal of Pragmatics, 43*(8),

2243-2253.

Nam, S. (2005). N-ary quantifiers and the expressive power of DP-compositions. *Research on Language and Computation, 3*(4), 411-428.

Noh, E. (2009). The Korean conditional markers *myen* and *tamyen*: Epistemicity vs. modes of language use. *Journal of East Asian Linguistics, 18*(1), 21-39.

Noh, E., Choo, H., & Koh, S. (2013). Processing metalinguistic negation: Evidence from eye-tracking experiments. *Journal of Pragmatics, 57*(1), 1-18.

Oh, S. (2010). Invoking categories through co-present person reference: The case of Korean conversation. *Journal of Pragmatics, 42*(5), 1219-1242.

국문초록

영어 원어민과 비원어민의
언어학 학술논문 어휘묶음 비교

최빛나
서울대학교 대학원
영어영문학과 대학원 전공

본 논문의 목적은 코퍼스를 기반으로 영어 원어민과 비원어민의 언어학 학술논문에서의 어휘묶음의 빈도, 구조, 기능을 연구하는 것이다. *Antconc 3.4.3*(Anthony, 2014)를 사용하여 4단어로 이루어진 어휘묶음을 91개의 학술 논문으로 구성된 909,259 단어의 코퍼스에서 추출하였다. 이 코퍼스는 이론 언어학 분야의 주요 저널에서 최근 11년 간 게재 된 논문들로 구성되었으며, 52개의 영어 원어민 저자의 논문과 39개의 한국어 원어민 저자의 논문으로 이루어져 있다. 어휘묶음 추출 시 빈도 및 분산 기준이 사용 되었고, 추출된 후에는 기존연구에 사용된 제거 기준을 적용하여 최종 어휘묶음 목록이 완성되었다. 분석 시에는 각각의 집단에서 가장 많이 쓰이는 20개의 어휘묶음이 비교 되었다. 구조적, 기능적 특질은 각각 Biber et al.(1999)과 Hyland(2008a)의 수정된 틀을 콘코던스(concordance) 분석을 통해 살펴 보았다.

이러한 양적, 질적 분석의 혼합 방법론의 사용을 통해 본 연구는 비원어민의 어휘묶음 과다사용을 발견했는데, 이는 비원어민의 메타담화(metadiscursive) 어휘묶음의 빈번한 사용 때문인 것으로 확인 되었다. 구조적으로는 양 집단 모두 전치사 기반, 동사 기반, 명사 기반 어휘묶음 순으로 선호하는 비슷한 양상을 보였지만, 비원어민은 수동태를 더 선호하는 모습을 보였다. 이는 비원어민의 경우 Hyland(2008b)가 말한 저자 정체성(authorial identity)을 드러내길 꺼려 하는 데서 기인하는 것으로 추정 된다. 기능 면에서 원어민과 비원어민은 전반적으로 동일한 양상을 보였는데, 텍스트 중심, 연구 중심, 참여자 중심 순으로 어휘묶음을 선호하는 것이 그러했다. 그러나 구체적인 영역에서는 차이를 보였는데, 먼저 저자의 개인 의견을 드러낼 때 원어민은 간접적 방식을 택하는 반면, 비원어민의 경우 보다 직접적인 표현인 *I/we*를 주로 사용했다. 이러한 비원어민의 자기언급(self-mention)은 이들에게 서술적 형용사가 익숙하지 않을 수 있다는 점을 암시한다. 기존연구와 달리 본 연구에서는 비원어민이 다양한 종류

의 헤지(hedge)를 사용하는 것으로 드러났는데, 이는 단순히 수만 많은 것이 아니라 Biber(2006)가 말한 확실성의 정도가 높은 것에서 낮은 것까지 다양한 종류를 구사하고 있는 것으로 밝혀졌다. 과거 “원어민 어휘묶음”(Hyland, 2000b)이라 불리던 이 어휘묶음을 비원어민이 많이 사용한 것은 이들의 전문성을 드러내는 반면, 높은 비율의 법조 동사(modal verb) 화행 동사(speech verb) 사용은 비원어민적 배경을 보여 주었다.

본 논문은 기존연구에서 보여 준 원어민과 비원어민의 어휘묶음 사용에서의 차이를 확인한 반면, 원어민과 비원어민 저자들의 어휘묶음 사용에 있어 새로운 공통점을 보여줌으로써, 원어민의 여부 뿐 아니라 전문성 역시 중요한 요소라는 점을 시사한다.

주요어 : 코퍼스, 어휘묶음, 영어 원어민, 비원어민, 언어학, 학술 논문
학번 : 2011-23083