The purpose of the study is to examine the discourse functions of lexical bundles in written university registers. Lexical bundles which are frequently found in written academic discourse are examined in order to analyze their functions in this register, comparing the frequency of each function. As the written register, theses of 72 EFL students were randomly chosen. The categorization was constructed based on the functional framework of lexical bundles in discourse offered by Biber et al. Every 4-word sequence in the theses was identified. The most frequent 4-word bundles in the theses were categorized and discussed accordingly.

Key words: Academic setting, EFL learners, lexical bundles, discourse functions, written discourse.

INTRODUCTION

The term ‘lexical bundle’ comes from the field of corpus linguistics. It was defined as “the most frequent recurring lexical sequences which can be regarded as extended collocations: sequences of three or more words that show a statistical tendency to co-occur (e.g., in the case of the)” (Biber and Conrad, 1999). The concept of lexical bundles goes back to Salem (1987) and the research he carried out on a corpus of French government texts. Butler (1997) and Altenberg (1998) subsequently employed the notion in their investigations based on Spanish and English corpora. However, the term ‘lexical bundle’ first explicitly appeared in the Longman Grammar of Spoken and Written English (Biber and Conrad, 1999), a monumental work entirely based on the British National Corpus of 100 million words.

In addition to Biber and Conrad’s term, these multi-word items have been studied under different terms such as fixed expressions (Moon, 1998), lexical phrases (Nattinger and DeCarrico, 1992), pre-fabs, readymade units, routines, formulas, using different criteria to define and identify multi-word items.

Although these words were given different names and they were investigated in terms of structures, their use in discourse came to the fore much later. They started to be seen as an especially productive approach in describing discourse in different contexts like university registers, academic registers (Biber and Conrad, 1999; Biber, and Barbieri, 2007; Hyland, 2008) or the language of journalistic prose (Cowie, 1992). The concept of lexical bundles has been used in several later studies (Biber et al., 2004; Cortes, 2004) to investigate common multi-word items in discourse. Different approaches have been
Table 1. Discourse functions of lexical bundles based on Biber et al. (2004).

<table>
<thead>
<tr>
<th>Stance Bundles</th>
<th>Discourse Organizers:</th>
<th>Referential Bundles:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epistemic stance bundles</td>
<td>Topic introduction bundles</td>
<td>Imprecision bundles</td>
</tr>
<tr>
<td>Desire bundles</td>
<td>Topic elaboration, clarification bundles</td>
<td>Bundles specifying attributes</td>
</tr>
<tr>
<td>Obligation (directive) bundles</td>
<td>Identification/focus bundles</td>
<td>Time/place/text-deixis bundles</td>
</tr>
<tr>
<td>Intention/prediction bundles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability bundles</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

developed, using different criteria for the identification of multi-word sequences. For example, some studies describe multi-word sequences that are idiomatic (e.g. expressions like *in a nutshell*), while other studies focus on sequences that are non-idiomatic but perceptually salient (e.g. *you’re never going to believe this*).

Moreover, different modes have also been used to examine these multi-word sequences. There has been considerable attention given to the ways in which formulaic language can be studied in spoken and written discourse (Biber et al., 2004; Cowie, 1992; Moon, 1998). These results were in contrast to previous analyses which regarded the use of prefabs a characteristic of spoken registers (Pawley and Syder, 1983). Lexical bundles were noted to be present in written and spoken registers alike and they were considered “basic building blocks for constructing spoken and written discourse” (Biber and Conrad, 1999, p.188). Moreover, further research found that in certain written registers lexical bundles are surprisingly common (Biber and Barbieri, 2007).

As shown in the studies mentioned above, lexical bundles are important building blocks in discourse. By developing a detailed taxonomy, Biber et al. (2004) found that the three main functions lexical bundles serve in discourse include expressing stance, organizing discourse and referential expressions. They provide a kind of pragmatic ‘head’ for larger phrases and clauses, where they function as discourse frames for the expression of new information. In this way, lexical bundles provide interpretive frames for the developing discourse (Biber and Barbieri, 2007).

Stance bundles express attitude or assessment, discourse organizers reflect the relationships between different parts of texts, and referential expressions refer to physical or abstract entities, or to other textual parts. Biber et al. (2004) noted that each of these main categories has several sub-categories which are associated with more specific discourse functions. The present study focuses on each of these main functions to examine their use in written university registers. Lexical bundles which are frequently found in written academic language are examined in order to analyze their functions in this register, comparing the frequency of each function.

Discourse functions of lexical bundles

Lexical bundles establish various discourse functions and they differ among themselves in functional characteristics. These functions are provided in Table 1.

As can be seen from Table 1, main functions of lexical bundles are stance bundles, discourse organizers, and referential bundles. In the following sections, each function will be further discussed.

Stance bundles

Stance bundles, the first major category, express attitudes or assessments of certainty that frame some other proposition. There are five functional sub-categories of stance bundles: epistemic bundles, desire bundles, obligation (directive) bundles, intention/prediction bundles, and ability bundles:

Epistemic stance bundles

These bundles express some degree of certainty.

It is possible that those grammatical structures which are most significant for students’ writing development at any given stage can only be explored productively at a level below explicit awareness.

Desire bundles

These are the bundles including certain words expressing hope or wish such as want, expect, etc..

I don’t want to deliver bad news to her.
I want you to take out a piece of paper and jot some notes down . . .
Obligation (directive) bundles

These are the bundles including modals of obligation and necessity such as have to, must, etc.

All you have to do is work on it.

Intention/prediction bundles

These are the bundles including modals of intention and prediction such as be going to, will, etc.

right now what we're going to take a look at are ones that are [...] positive and beneficial.

Ability bundles

These are the bundles including modals of ability such as be able to, can, etc.

I want you to be able to name and define those four curriculum category.

Discourse organizers

Discourse organizing bundles, the second major category, are used to indicate the overall discourse structure and to signal the informational status of statements. Discourse organizing bundles serve three functional sub-categories: topic introduction, topic elaboration/clarification, and referential identification/focus:

Topic introduction bundles

Such kind of bundles serves for introducing a new topic.

What I want to do is quickly run through the exercise . . .

Topic elaboration/clarification bundles

These bundles help one exemplify or clarify a topic.

It has to do with the start talks, with the Russians.

Identification/focus bundles

Such bundles help one identify or focus on a certain topic.

For those of you who came late I have the, uh, the quiz.

Referential bundles

The third major category is referential bundles that generally identify an entity or single out some particular attribute of an entity as especially important. There are three functional sub-categories included under referential bundles: imprecision bundles, bundles specifying attributes, and time/place/text-deixis bundles:

Imprecision bundles

These bundles refer to something in a way which is not clear or exact.

I think really we now have what about, six weeks left in class or something like that.

Bundles specifying attributes

These bundles are in attributive position, usually including quantifiers.

It creates a little bit of wealth.

These figures give an idea of the size of the ethnological community in Russia.

...students must define and constantly refine the nature of the problem . . .

Time/place/text-deixis bundles

Such bundles refer to the concept of time, place or text, usually including prepositions.

Children in the United States are not formally employed in farm work . . .

She's in that... uh... office down there... at the end of the hall... . . .

As shown in Figure 4.4, . . .

Lexical bundles in literature

Results of corpus-driven research motivated a growing number of studies that explored different structural types of recurrent multiword chunks in various kinds of written corpora: recurrent word combinations (Altenberg, 1993), prefabricated patterns (Granger, 1998), phrasal lexemes (Moon, 1998), highly recurrent word combinations (De Cock, 2000), and lexical bundles (Biber and Conrad, 1999; Biber et al., 1999), all of which are identified in a language on the basis of their frequency of co-occurrence.

Biber (1999), in his study, compared lexical bundles in conversation and academic prose, while Cortes et al.
Cortes (2002) compared the use of lexical bundles in the corpus of freshman writing, academic prose and conversation in her study in Freshman Composition. She inferred that the main function of lexical bundles is establishing discourse relationship which helps students to speak and write with greater fluency and accelerating language acquisition process.

Hyland (2008) explored the forms, structures and functions of 4-word bundles in a 3.5 million word corpus of research articles, doctoral dissertations and Master’s theses in four disciplines. He indicated that the common use of lexical bundles such as it has been noted that in academic written genres helps to signal the text register to readers and reduce processing time by using familiar patterns to link elements of new information. Thus, text receivers are able to sort out what is natural from what is merely grammatical and judge whether a particular bundle ‘sounds right’ in that context. Therefore, as can be seen is a frequent and unremarkable bundle in academic writing while the equally possible as you can see or as can be observed are rarely encountered. Analyzing a range of written corpora, other researchers suggested that, in addition to already established classes of formulaic sequences (sayings, proverbs, speech formulae, and idioms), a large number of language units were found to cooccur in preferred order without being governed by specific grammar rules (Altenberg, 1993; Biber and Conrad, 1999; Granger, 1998; Moon, 1998; Sinclair, 1999; Wray, 2002).

In their study Biber and Barbieri (2007) investigated the use of lexical bundles in written registers and found that lexical bundles are very common in written course management. They included both instructional registers and student advising/management registers such as office hours, class management talk, written syllabi, etc. They suggested that both course management and institutional writing use a large number of lexical bundle types. However, the functional distribution of bundles was strikingly different in these two registers. In written course management, over half of all bundle types were stance bundles. Referential bundles were relatively common in that register. On the other hand, in institutional writing, almost 70% of all bundle types were referential; stance bundles and discourse organizers were considerably less common in that register.

### Research questions

As an exploration of the discourse functions of lexical bundles in written registers, this study addresses the following four research questions:

1. Do all of the three functions exist in the written register?
2. If so, among the three main discourse functions of lexical bundles, which is the most frequent one in written registers?
3. Which sub-category is the most frequent one among the three main discourse functions?
4. What are the most frequent lexical bundles creating each main discourse function?

### METHODOLOGY

The present study is parallel with these earlier studies in that it also investigates the frequency of these bundles. However, it goes one step further by examining the frequencies of bundles with respect to their discourse functions and exploring the most common ones. More specifically, the study aims to examine a written register - theses of 72 EFL students who submitted their theses as a final project required in the course “research methods”.

With respect to participants, all of them were from the Foreign Language Teaching Department of the university. All of the participants had two years of education in EFL department. Among them, 44 were females and 28 were males. Their ages ranged from 19 to 22 (mean: 20.5).

As for data collection, in order to investigate lexical bundles in one disciplinary domain, educational sciences, 10 theses were selected from EFL Department. All studies were submitted within the period between 2013-2014. First, theses were collected and their introduction and conclusion parts were taken into account. The categorization was constructed based on the functional framework of lexical bundles in discourse mentioned earlier (Biber et al., 2004). Every 4-word sequence in the theses was identified.

When it comes to data analysis, lexical bundles were identified by two scorers. One of them was the researcher and the other one was another instructor. This determined the quantitative mode of the research. To limit the scope of the investigation, only recurrent sequences of 4-word bundles were analyzed in detail. They have been classified according to their function. Then a frequency count was applied for the comparison of lexical bundles in the theses of the selected disciplinary domain; first main categorization was formed. Then, main categorization was followed by sub-categorization. The corresponding function of each bundle was assigned by two scorers. Both scorers used the categorization of Biber et al (2004) for coding the data. They exhibited a 81% agreement rate in the categorization of bundles. Then, they came together to come to an agreement for the ones previously not agreed.

### RESULTS

#### Distribution of lexical bundles

Table 2 displays the distribution of bundle types in the written register according to their frequencies. While categorizing, their sub-functions are not considered. Instead, the categorization was formed according to their frequencies.

Table 2 highlights the fact that many of the most frequent 4-word bundles in academic writing are very
Table 2. Most frequent 4-word bundles in the theses.

<table>
<thead>
<tr>
<th>Stance bundles</th>
<th>Freq.</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>can be used to</td>
<td>3</td>
<td>2.5</td>
</tr>
<tr>
<td>it is possible that</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td>was found to be</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td>is likely to be</td>
<td>2</td>
<td>1.6</td>
</tr>
</tbody>
</table>

Discourse Organizers
one of the most                       | 10    | 8.4 |
the results of the                     | 9     | 7.6 |
on the other hand                      | 6     | 5   |
the fact that the                      | 6     | 5   |
as a result of                         | 5     | 4.2 |
with respect to the                    | 3     | 2.5 |
as well as the                         | 2     | 1.6 |
to that of the                         | 2     | 1.6 |
in the form of                         | 2     | 1.6 |
it is important to                      | 2     | 1.6 |
these results suggest that             | 2     | 1.6 |

Referential Bundles
was one of the                         | 12    | 10.1|
as part of the                         | 7     | 5.9 |
the number of the                      | 5     | 4.2 |
the aim of the                         | 5     | 4.2 |
at the end of                          | 5     | 4.2 |
the role of the                        | 4     | 3.3 |
the end of each                        | 3     | 2.5 |
at the beginning of                    | 3     | 2.5 |
in the present study                   | 2     | 1.6 |
to the extent that                     | 2     | 1.6 |
in the case of                         | 2     | 1.6 |
in each of the                         | 2     | 1.6 |
the nature of the                      | 2     | 1.6 |
the relationship between the           | 2     | 1.6 |
at the time of                         | 2     | 1.6 |
on the basis of                        | 2     | 1.6 |

It can be seen from Table 3 that in the first main category—stance bundles, the only sub-category in which some related lexical bundles with discourse functions are found is "epistemic bundles". On the other hand, the second main category, discourse organizers, includes three sub-categories for all of which certain discourse functions were detected. The final main category, referential bundles, has only one sub-category—imprecision bundles, for which no discourse function was found. Each category along with its sub-categories will be discussed separately in the next section.

Stance expressions

Table 4 shows that among the sub-categories of stance expressions, epistemic stance bundles are the only detected ones. Other sub-categories have not been detected in the theses.

Among the stance bundles, can be used to is the most common one (2.5%).

Discourse organizers

Table 5 reflects that topic introduction bundles and identification/focus bundles are the frequent ones. Topic elaboration/clarification bundles are not common so much.

Among the discourse organizers, one of the most is the most frequent one (8.4%).

Referential expressions

Table 6 highlights that bundles specifying attributes are the most common one (11). It is followed by time/place/text-deixis bundles (5). Imprecision bundles have not been detected in the theses.

Among the referential bundles, was one of the is the most common one (10.1%).

DISCUSSION

The first research question investigated whether all of the three functions exist in the written register. The results reflected in Table 3 showed that all of the three functions exist to some extent.

The next question asked for the most frequent main function in written register. The results revealed that referential bundles are the most frequent main function (16). It is followed by discourse organizing bundles (11) and stance bundles (4).

The third research question went one step further by...
Table 3. Overall frequency of each main categories and their sub-functions.

<table>
<thead>
<tr>
<th>Stance Bundles:</th>
<th>freq</th>
<th>(7.6%)</th>
<th>Discourse Organizers:</th>
<th>freq</th>
<th>(41.5%)</th>
<th>Referential Bundles:</th>
<th>freq</th>
<th>(50.8%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epistemic stance bundles</td>
<td>(9)</td>
<td>0.09</td>
<td>Topic introduction bundles</td>
<td>(25)</td>
<td>0.41</td>
<td>Imprecision bundles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Desire bundles</td>
<td>(0)</td>
<td>0.00</td>
<td>Topic elaboration, clarification bundles</td>
<td>(22)</td>
<td>0.22</td>
<td>Bundles specifying attributes</td>
<td>(45)</td>
<td>0.45</td>
</tr>
<tr>
<td>Obligation (directive) bundles</td>
<td>(0)</td>
<td>0.00</td>
<td>Identification/focus bundles</td>
<td>(22)</td>
<td>0.22</td>
<td>Time/place/text-deixis bundles</td>
<td>(15)</td>
<td>0.15</td>
</tr>
<tr>
<td>Intention/prediction bundles</td>
<td>(0)</td>
<td>0.00</td>
<td>Ability bundles</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Overall frequency of each sub-category related to stance expressions.

<table>
<thead>
<tr>
<th>Stance Bundles:</th>
<th>freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epistemic stance bundles</td>
<td>(9)</td>
</tr>
<tr>
<td>Desire bundles</td>
<td>(0)</td>
</tr>
<tr>
<td>Obligation (directive) bundles</td>
<td>(0)</td>
</tr>
<tr>
<td>Intention/prediction bundles</td>
<td>(0)</td>
</tr>
<tr>
<td>Ability bundles</td>
<td>(0)</td>
</tr>
</tbody>
</table>

Table 5. Overall frequency of each sub-category related to discourse organizers

<table>
<thead>
<tr>
<th>Discourse Organizers:</th>
<th>freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic introduction bundles</td>
<td>(25)</td>
</tr>
<tr>
<td>Topic elaboration, clarification bundles</td>
<td>(2)</td>
</tr>
<tr>
<td>Identification/focus bundles</td>
<td>(22)</td>
</tr>
</tbody>
</table>

Table 6. Overall frequency of each sub-category related to discourse organizers

<table>
<thead>
<tr>
<th>Referential freq</th>
<th>Bundles:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imprecision (0)</td>
<td>bundles</td>
</tr>
<tr>
<td>Bundles (45)</td>
<td>attributes</td>
</tr>
<tr>
<td>Time/place/text-deixis (15)</td>
<td>bundles</td>
</tr>
</tbody>
</table>

asking the most common sub-category. The results suggested that among the sub-categories of stance expressions, epistemic stance bundles are the most frequent ones. With respect to discourse organizers, topic introduction bundles and identification/focus bundles are the frequent ones. Finally, in referential expressions, bundles specifying attributes are the most common one.

The last research question was concerned with the most frequent lexical bundles creating each main discourse function. As for stance expressions, can be used to (3) is the most frequent one. With respect to discourse organizers, one of the most (6) is the common one. In the
last category - referential expressions, was one of the (7) is the frequent bundle.

Conflict of Interests

The author has not declared any conflict of interests.

REFERENCES


