Abstract

This paper reviews research into the role of lexical bundles, multi-word expressions, that help to shape meanings and contribute to coherence in a text. The use of a wide range of extended collocations helps writers to achieve naturalness in their language use. Based on the corpora method, statistical regularities, variations in the forms and functions of bundles, differences in use by PhD theses and published articles have been provided. Findings refer to the frequency, structures, distribution, and variations by discipline in bundle use. Lexical bundles play an essential role in clarifying what is expressed and help nonnative writers establish a credible academic voice in their discipline.

Keywords
lexical bundles, corpus analysis, coherence, frequency, disciplinary variations

Introduction

Multi-word expressions, referred to as "bundles", "clusters" or "chunks", help to identify a text as belonging to a register, for instance, either conversational, journalistic, legalese, or academic. Use of these bundles signal a writer’s language competence and fluency, while absence of such clusters may reveal lack of fluency of a novice or newcomer to that discipline community [1]. Has well [2] suggests:

there can be little doubt that as writers mature they rely more and more on collocations and that the lesser use of them accounts for some characteristic behaviour of apprentice writers. Gaining control of a new register therefore requires a sensitivity to experts’ preferences for certain sequences of words over others that might be equally possible.

The wide use of formulaic language helps writers to achieve naturalness in language use. The findings indicate that collocations that "are divergent from native speaker norms, take longer to process when reading” [3].

According to Biber, Conrad and Cortes, these multi-word expressions in academic prose often serve to bridge two phrases, either two noun phrases or a noun and a verb phrase [4]. In all of these roles, they function like scaffolding for new information [5]. Biber categorized these clusters under three essential functional roles: stance expressions, discourse organizers, and referential expressions [6].

Research has shown that learners of English from other language groups than English produce language features in their writing which differs from native speaker norms. Granger has shown that in many cases writers overuse, underuse or misuse particular language functions and exponents, such as linking and stance adverbs, in ways divergent from native speaker writers [7]. According to the author’s experience, Estonian writers tend to overuse lexical bundles that express hedging, for instance: It could be calculated, etc. Allen’s studies focused on the degree of convergence between writers’ production and published and/or native speaker writing [8].

On the basis of findings derived from corpora analysis, focus will be on how novice writers can establish their voice in the discipline community by composing texts that sound natural. Different categories of lexical bundles, their frequency of occurrence and variations across disciplines are described.

1 The Academic Voice

The awareness of lexical bundles can empower nonnative writers who face the challenge of not finding the right words to express their ideas, but finding the best way to present those ideas to give the right impression. Science writing involves creating a so-called "discoursal self" [9] or a "professionally acceptable persona" [10].

This discoursal self is called the writer’s voice that refers to the rhetorical forms and features that a writer chooses to make use of in creating a desired impression on the reader. Readers form an opinion of the writers’ credibility according to their perception of that voice. Thus, by language use that carries credibility a writer becomes a member of a disciplinary community.

Use of lexical bundles in appropriate register helps to create the writer’s voice. For instance, the second personal pronoun is not appropriate in the academic context:

What you need to know is that qZSIs suit for different renewable power applications.

Instead, an impersonal lexical bundle is appropriate: It is important to know that qZSIs suit for different renewable power applications.

Another alternative can be: As we have seen, qZSIs suit for different renewable power applications.
Here the use of *we* according to the analysis of academic corpus has revealed that the "inclusive *we* is heavily used to bind writer and reader together as members of a disciplinary in-group," while the use of the second person *you* and *your* is rare [10].

Nonnative writers encounter certain difficulties in the usage of these multi-word sequences that help to establish credible English academic voice in their writing and speaking. So, they need to be made aware of how lexical bundles are used, in particular, within their discipline areas.

### 2 Categorization of lexical bundles

Lexical bundles are categorized both structurally, in terms of grammatical types, and functionally, according to their meanings in texts. Biber categorized lexical bundles under three functional roles: stance expressions, discourse organizers, and referential expressions [6].

Examples taken from IEEE Transactions:

Stance expressions distinguish what is factual, what is likely, and what is doubtful –

*It is true that* power electronics researchers have tried to reduce the cost and complexity of having two separate switch-mode converters.

Discourse organizers introduce new topics or signal that the topic will be elaborated on or clarified or that a point of contract will be introduced –

*It is estimated that* the efficiency will grow by 10%.

Referential expressions refer to something that can be concrete or abstract in order to highlight it or to point out a particular attribute –

**Table 1.** Structural patterns of 4-word clusters in 4 disciplines [1, p. 48]

<table>
<thead>
<tr>
<th>Structure</th>
<th>Examples</th>
<th>% of all structures</th>
<th>% of all cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>prepositional phrases</td>
<td>on the other hand, at the same time, in the present study, with respect to the at the end of, the nature of the, the beginning of the, a large number of in the case of, at the end of, as a result of, on the basis of, in the context of the fact that the, one of the most, the extent to which, the relationship between the is shown in figure, is based on the, is defined as the, can be found in it is important to, it is possible that, it is difficult to, it was found that may be due to, is a matter of, is due to the, be the result of as shown in figure, should be noted that, is likely to be as well as the</td>
<td></td>
<td></td>
</tr>
<tr>
<td>noun phrase + <em>of</em></td>
<td></td>
<td>22</td>
<td>27</td>
</tr>
<tr>
<td>prepositional phrase + <em>of</em></td>
<td></td>
<td>22</td>
<td>19</td>
</tr>
<tr>
<td>noun phrases with other post-modification</td>
<td></td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>passive + prepositional phrase fragment</td>
<td></td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>anticipatory <em>it</em> + verb/adj</td>
<td></td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>verb (<em>be</em>) + noun phrase</td>
<td></td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

In the case of conventional use the bus voltage would be 400 V.

Commonly, grammatical accuracy in lexical bundles is high. However, examples of misuse of the bundles surrounding the stem *result of* have been found [8].

Examples:

*The result of this experiment was expressed by the following graphs.*

Experiments tend to produce more than one result, so it is appropriate to write *the results of the experiments.*

The reason is that in other expressions the singular form is used:

*(X) (to be) *a/the result of Y

*As a result of (X), (Y)*

These forms are highly frequent bundles found in scientific discourse.

In the above examples it is evident that although many of these units of words can be eliminated from the sentence with the sentence still remaining grammatical, each lexical bundle unit plays an essential role in clarifying what is expressed. Without them, the sentence may be confusing or the significance of what is presented lost to the reader.

Hyland analyzed a corpus of 3.5 million words and found 130 different 4-word clusters in 4 disciplines, including electrical engineering. Table 1 shows the results of his study based on structural classification.
According to Hyland, on the other hand was the most frequent cluster, occurring 100 times per million words, and was over twice as common as the next-placed clusters, at the same time and in the case of. As can be seen, most of the clusters in academic writing are parts of noun or prepositional phrases and end with prepositions, and articles. Noun phrases and it phrases together comprise 70% of 4-word patterns in academic discourse. It is interesting to note that some researchers have suggested that as much as 80% of natural language could be patterned in this way [11].

3 Disciplinary variations in the frequency and structure of lexical bundles

In his corpora analysis, Hyland has also explored frequencies and structures of 4-word disciplinary bundles and has found interesting disciplinary differences [12]. Out of texts in 4 different disciplines: electrical engineering, business studies, applied linguistics, and biology, electrical engineering texts were found to contain the greatest range of bundles (213 different bundles), as shown by Table 2 below. Biology, on the other hand, has the smallest range of bundles. Many bundles found in engineering were not found in other disciplines and engineering writers rely more on pre-fabricated structures than writers in other fields. It could be a consequence of the relatively abstract and graphical nature of engineering communication. Typically, engineering rhetoric depends more on visual and numeric representation.

In addition to different frequencies, structures of bundles were also found to be different across disciplines. As can be seen in Table 3, the noun phrase with of- is the most common structure phrase in each discipline.

Table 2. Bundle frequency information [12]

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Different bundles</th>
<th>Total cases</th>
<th>% of total bundle words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electrical engineering</td>
<td>213</td>
<td>4562</td>
<td>3.5</td>
</tr>
<tr>
<td>Business studies</td>
<td>144</td>
<td>3728</td>
<td>2.2</td>
</tr>
<tr>
<td>Applied linguistics</td>
<td>141</td>
<td>4631</td>
<td>1.9</td>
</tr>
<tr>
<td>Biology</td>
<td>131</td>
<td>2909</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Table 3. Main structures of bundles across disciplines (%)

<table>
<thead>
<tr>
<th>Structure</th>
<th>Biology</th>
<th>Electrical</th>
<th>Applied</th>
<th>Business</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun phrase + of</td>
<td>23.7</td>
<td>22.3</td>
<td>22.9</td>
<td>28.5</td>
<td>24.4</td>
</tr>
<tr>
<td>Passive + prep.</td>
<td>31.3</td>
<td>29.8</td>
<td>6.9</td>
<td>9.0</td>
<td>19.3</td>
</tr>
<tr>
<td>Other prep. phrase</td>
<td>13.7</td>
<td>11.6</td>
<td>24.4</td>
<td>19.7</td>
<td>17.5</td>
</tr>
<tr>
<td>Prep. phrase + of</td>
<td>9.2</td>
<td>7.9</td>
<td>19.9</td>
<td>16.0</td>
<td>13.5</td>
</tr>
<tr>
<td>Noun phrase + other modification</td>
<td>9.4</td>
<td>10.8</td>
<td>9.6</td>
<td>12.4</td>
<td>10.6</td>
</tr>
<tr>
<td>Others</td>
<td>6.4</td>
<td>9.2</td>
<td>10.7</td>
<td>9.9</td>
<td>9.5</td>
</tr>
<tr>
<td>Anticipatory it</td>
<td>6.3</td>
<td>8.4</td>
<td>5.6</td>
<td>4.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Totals</td>
<td>100</td>
<td>100</td>
<td>1000</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

The engineering texts employ significantly more passive bundles, followed by a prepositional phrase marking location or logical relation.

Examples:

The experiment setup is shown in Fig. 4.
The value of Xi is given by Eq. (4).
The measurement is based on the evaluation of infrared images produced by thermal waves.

Thus, interesting disciplinary differences are demonstrated by the frequency of occurrence and structure of lexical bundles. According to corpora-based research, many lexical bundles found in engineering are specifically disciplinary. Another characteristic is that engineering writers rely more on prefabricated structures that could be a consequence of the relatively abstract and graphical nature of engineering communication. In addition, engineering rhetoric depends more on visual and numeric representation.

Conclusions

Though statistical regularities of language use are for the analyst, they actually reflect a reality for users. Findings of corpora analyses indicate to the essential role of lexical bundles in science writing and enable writers expand their academic rhetorical toolkit by use of the multi-word lexical bundles. Nonnative writers produce language features in their writing which differs from native speaker norms. Usage of these expressions helps nonnative and novice writers establish credible English academic voice in their discipline. The lexical bundles are intentionally used by professional academic writers to establish solidarity and distinguish themselves as proficient writers. Thus it is essential to raise a research writer’s awareness of the ability of a text reader to sort out what is natural from what is merely grammatical and judge whether a collocation is appropriate.
References


