



Potential of L1 and L2 Corpora to Identify Target Lexical Bundles for Argumentative Essay Writing

Ryo Sawaguchi
(Kansai University)

Sawaguchi, R. (2024). Potential of L1 and L2 corpora to identify target lexical bundles for argumentative essay writing. *Asia Pacific Journal of Corpus Research*, 5(1), 1-21.

This study aimed to identify *target* lexical bundles (e.g., *on the other hand*, *at the same time*) for argumentative essay writing and rank them in order of teaching priority for Japanese learners. Despite significant functional roles of lexical bundles in academic writing, the inclusion of lexical bundles in argumentative writing had been underexplored. Since argumentative writing skills help undergraduate students prepare for their academic careers (e.g., writing papers), the lexical bundles under this genre deserve more attention. This study first extracted 78 target bundles from L1 argumentative essay corpora (International Corpus Network of Asian Learners of English: ICNALE and Louvain Corpus of Native English Essays: LOCNESS). The study then classified the bundles according to their discourse functions and semantic transparency to estimate the learnability for Japanese learners in L2 compatible corpora with the ICNALE. The results showed that learners had difficulty using the bundles with referential functions (e.g., *in the form of*) and semantic opaqueness (e.g., *when it comes to*), suggesting that the bundles in these two categories should be prioritized among the 78 bundles.

Keywords: Corpus Linguistics, English for Academic Purposes, Formulaic Language, Lexical Bundles, Argumentative Writing

1. Introduction

The interest in recurring word clusters (e.g., *on the other hand*, *at the same time*) has been increasing. These word clusters are called formulaic language (Laufer, 2021; Schmitt, 2004; Siyanova-Chanturia & Pelicer-Sánchez, 2019; Wray, 2002) or multi-word expressions (Hyland, 2008; Martinez, 2013). Wray (2002) theorized formulaic language, arguing that it comprises predictable, fixed, or semifixed chunks stored in a speaker's mental lexicon. These features of formulaic language allow native (L1) English speakers to produce language with little cognitive burden, achieving naturalness (Pawley & Syder, 1983).

Formulaic language not only strengthens the effectiveness of L1 English speakers' language use but also benefits English learners (Boers & Lindstromberg, 2009; Conklin & Schmitt, 2008; Khodadady & Shamsaee, 2012). Formulaic language is especially important for learners to succeed in English for academic purposes (EAP) situations. Among the types of formulaic language in EAP, *lexical bundles* are often targets of investigation. Lexical bundles are recurring clusters of words (e.g., *what I mean*, *as a result of*) that occur frequently in speech or writing (Granger, 2019). Thus, lexical bundles are commonly observed in different academic discourse, and learners should familiarize themselves with the lexical bundles specific to each register. Regarding the receptive (reading/listening) aspect of lexical bundles, as described by Biber et al. (2004), university students manage lexical bundles that are ubiquitous in classroom teaching and materials such as textbooks. Additionally, university students learning English as a second language (L2) in English-speaking countries must acquire listening

comprehension of lexical bundles in academic lectures (Liu & Chen, 2020). Regarding productive knowledge, especially in writing, learners should have a good command of lexical bundles to engage in the academic world. In other words, they need to acquire accepted expressions in academia to publish their articles (Shamsabadi et al., 2017).

The aforementioned examples highlight the importance of lexical bundles in many aspects of EAP. One area that deserves increased attention, however, is how to teach college-level writing genres such as essays and what learning/teaching materials would be useful for that genre.

Essays are a key pedagogical process genre in that they facilitate students' critical thinking skills to articulate and support their position (Nesi et al., 2017). Specifically, the most common and basic academic writing genre for undergraduate students is the argumentative essay (Wu, 2006). This type of writing is often involved in the discussions that aim to gain insights about the lexical bundles that could improve students' basic academic writing skills (Appel & Murray, 2020; Granger, 2017; Nam, 2017). This information implies that argumentative essay writing has pedagogical potential to inform English for general academic purposes (EGAP) practice. Thus, identifying "target" lexical bundles is essential and would promote university students developing their academic writing skills in the form of argumentative essay writing. This study aims to achieve this goal by identifying statistically important lexical bundles and the factors affecting their learnability, namely, functions, frequency, and semantic transparency.

2. Literature Review

2.1. Lexical Bundles and Their Roles in Academic Writing

Lexical bundles (e.g., *as can be seen, if you look at*) in academic English have been investigated (Biber & Barbieri, 2007; Byrd & Coxhead, 2010; Cortes, 2004; Hyland, 2008). Lexical bundles differ from other types of formulaic language, including "collocations, e.g., *take place*; phrasal verbs, e.g., *let down*; idioms, e.g., *at the top of a hat*" (Laufer, 2021, p. 89). Lexical bundles build discourse connections; for instance, *on the other hand* indicates that the writer/speaker is transitioning to a different side of the topic. Thus, studies on academic English have focused on lexical bundles writers have employed to suit their communicative purposes in EAP. Biber and Barbieri (2007, p. 273) pointed out, "The extent to which a speaker or a writer relies on lexical bundles is strongly influenced by their communicative purposes." Other related studies have yielded valuable insights into the discourse functions of lexical bundles. Biber et al. (2004) conducted a systematic analysis of lexical bundles in academic English and established a three-category taxonomy: (1) stance expressions (e.g., *I don't think so*), (2) discourse organizers (e.g., *If you look at*), and (3) referential expressions (e.g., *and this is a*) (pp. 384–388). The authors observed a difference in the functions of lexical bundles between spoken and written registers of university language. Biber et al.'s (2004) taxonomy was later modified or expanded in similar studies (Hyland, 2008; Simpson-Vlach & Ellis, 2010) that investigated functions of lexical bundles in different academic registers.

Another crucial aspect of lexical bundles is that they are often genre-specific. Omidian et al. (2021) analyzed the lexical bundles in British Academic Written English (BAWE). BAWE collects genres of academic writing, such as essays and case study assignments, written by British university students. Omidian et al. (2021) found that a greater number of stance bundles (bundles that show a writer's position on the topic) were observed in argumentative writing genres such as essays and critical papers than in case studies. For instance, of the three genres, essays used evaluation (e.g., *it is obvious that*), a subcategory of the stance bundles, the most. The authors discussed that a reason for this finding might be that the argumentative writing genre requires evaluative and careful observation from different perspectives to present convincing arguments. Referring to Omidian et al.'s (2021)

results, Laufer (2021) suggested the need for genre-specific (e.g., argumentative) academic lexical bundles (which she termed *formulas*) to improve their suitability for different EAP educational purposes.

From this view, lexical bundles in argumentative essay writing seem to have potential for EAP practice. Especially in English for general academic (EGAP) contexts, in which undergraduate students gain basic academic skills, argumentative essay writing can be a key component genre. Johnson (2018) stated that rhetorical features in genres such as arguing and explaining cover a wide range of text types. For example, the argumentative writing genre includes research papers and argumentative essays because they share the rhetorical purpose of stating the author's position and supporting it with evidence. In this regard, university students should benefit from learning the argumentative writing genre to better prepare for using English for specific academic purposes in their disciplines. Gardner et al. (2018) described this as "progression routes leading students from more general to more discipline-specific writing," stating that "pre-university or first year composition teaching for multiple disciplines may tend towards EGAP" (Gardner et al., 2018, pp. 646–647). Therefore, undergraduate students should become familiar with argumentative essay writing if they are planning to pursue academic careers.

2.2. Functional Features of L2 Lexical Bundles

Owing to the importance of lexical bundles in EAP contexts, studies (Bychkovska & Lee, 2017; Chen & Baker, 2016; Juknevičienė, 2009; Nam, 2017; Staples et al., 2013) have investigated L2 students' acquisition of lexical bundles in academic writing, including argumentative essays. These studies are motivated by pedagogical goals to identify what lexical bundles learners can and cannot use, proposing suggestions for teaching. The studies achieve this by systematically analyzing the discourse functions of lexical bundles used by L2 learners. Juknevičienė (2009) and Nam (2017) have compared L2 English learners' use of lexical bundles with that of L1 English speakers. Applying Biber et al.'s (2004) functional analysis framework, they both found that the learners tended to rely on certain types of lexical bundles (e.g., stance bundles). This differs from L1 English speakers, whose bundle uses are dominated by noun-based referential bundles (e.g., *in the form of*) typical of academic writing (Biber et al., 2004; Chen & Baker, 2010; Simpson-Vlach & Ellis, 2010). Based on their findings, Juknevičienė (2009) and Nam (2017) have suggested that learners should be aware of standard academic writing styles that often involve referential bundles. From a developmental perspective of L2 learners' acquisition, Staples et al. (2013) and Chen and Baker's (2016) studies are notable: they used objective criteria such as TOEFL scores and the levels of the Common European Framework of Reference for Languages (CEFR: Council of Europe, 2001) to inquire about how learners at different proficiency levels acquire lexical bundles. For example, Chen and Baker (2016) found that learners use fewer quantifying (e.g., *there are a lot of*) expressions in referential bundles as they become more proficient; their writing styles tend to be similar to written (formal) English. The highest-proficient learners start to use *a great deal of* rather than *there a lot of*, which is more conversational than the former.

The aforementioned studies have provided valuable insights into the acquisition of lexical bundles. Pedagogical implications might be that there should be "target" functional categories (e.g., referential) of lexical bundles for learners at different proficiency levels. However, other criteria, such as semantic transparency, L1 congruency, and frequency, leave room for discussion. As Siyanova-Chanturia and Pelicer-Sánchez (2019) stated, semantic transparency is one of the factors that affect the learnability of formulaic language including lexical bundles. In the case of collocations, one of formulaic language, Sawaguchi & Mizumoto (2022) found that Japanese L2 English learners tend to use fewer delexical (semantically opaque) collocations such as *make decisions* than L1 English speakers. This implies the difficulty of the acquisition of semantically opaque formulaic language. The same could be true for lexical bundles. L1 congruency (Yamashita & Jiang, 2010) and frequency (Martinez, 2013) also matter.

In this regard, Martinez (2013) effectively integrated frequency and semantic transparency, creating what he termed the frequency-transparency framework (FTF) (p.190).

2.3. Semantic Transparency as a Factor of Learnability

Martinez (2013) defined semantic transparency as “how easy/difficult an expression is to interpret” (p. 187) from individual words. For example, inferring the meaning *happen* from an expression *take place* is difficult because neither *take* nor *place* is used in literal meaning. Thus, they are not semantically transparent. Martinez (2013) considered this semantic transparency along with frequency to rank multi-word expressions in the order of teaching/learning priority, creating the FTF. In this framework, less semantically transparent (opaque) but frequent expressions would be introduced first because learners may not understand their meanings despite the expressions being frequently used in English; by contrast, less frequent but semantically transparent expressions would receive lower priority than the former because they are easily understandable to learners and are used less than frequent expressions. Generally, lexical bundles are more semantically transparent (Siyanova-Chanturia & Pelicer-Sánchez, 2019) than other multi-word expressions. The bundles *as a result of* and *I am going to* are literal in meaning, and the idiom *beat around the bush* is not. For this reason, semantic transparency of lexical bundles has not received much attention in the literature. Nevertheless, not all lexical bundles are semantically transparent. As Sykes (2017) demonstrated, some lexical bundles can be semantically opaque. For example, when the bundle *if you look at* is used in *if you look at page fifty-five* (p. 50), *look at* is literal and semantically transparent. However, in the sentence *if you look at culture as a system of values* (p.51), *look at* carries a different meaning: a person analyzes an issue from a certain perspective. Moreover, some lexical bundles have idiomatic uses. The bundle *when it comes to* cannot be understood using the literal meaning.

Because of the role of semantic transparency in formulaic language learning burden (Martinez, 2013; Siyanova-Chanturia & Pelicer-Sánchez, 2019), considering semantic transparency in lexical bundles could lead to more precise or detailed inclusion of the lexical bundles to teach.

2.4. Summary of the Literature

In summary, the findings in the literature suggest two crucial factors for the inclusion of target lexical bundles:

- Functional categories (e.g., referential, stance, discourse bundles)
- Semantic transparency (e.g., *when it comes to* [opaque] vs. *as a result of* [transparent])

This study adopts the aforementioned two criteria to identify target lexical bundles for argumentative essay writing. Specifically, this study employs corpora of L1 English speakers and L2 English learners (Japanese). The study aims to fill the following gaps in the literature:

First, the functions of “target” bundles for argumentative essay writing and how learners use them require further research. Studies on learners’ acquisition of lexical bundles have proposed target-like bundles (frequently used by L1 English speakers) in argumentative writing. However, according to the author’s review of the literature, no study has defined the bundles as a “target” in that writing genre. An effective approach for easing the learning burden would be for learners to focus exclusively on items in target bundles that they are not familiar with. This study first extracts “target” bundles from an L1 corpus and then explores how learners in an L2 corpus use the bundles in different functions.

Second, according to the author’s review of the literature, no studies have applied the FTF to an analysis of L1 and L2 corpora. Martinez (2013) noted that the degree of transparency may be subject to L1 congruency. For instance, Spanish and English have an equivalent translation, *take place*, but

Portuguese does not (p. 188). In this case, the degree of transparency differs for Spanish and Portuguese learners of English. Learner corpora such as the International Corpus Network of Asian Learners of English (ICNALE; Ishikawa, 2023) have Japanese learner data. This makes observing how learners use semantically opaque/transparent expressions possible.

As Granger (2017) suggested, target items such as L1 wordlists should be complemented by an L2 corpus. For example, L2 corpus data provide information on the learnability of target items. Based on L1 and L2 corpora, the research questions of this study are as follows:

RQ1: How are target bundles functionally categorized, and what items should Japanese learners focus on?

RQ2: To what extent is the FTF applicable to target bundles?

This study is motivated by two hypotheses emerging from the literature: L2 learners have difficulty mastering lexical bundles with referential functions and semantic opaqueness.

3. Methods

3.1. Compilation of Target Bundles for Argumentative Essay Writing

To extract “target” bundles for argumentative essay writing, this study used two widely available argumentative L1 essay corpora: ICNALE (Ishikawa, 2023) and The Louvain Corpus of Native English Essays (LOCNESS) (Granger, 1998). These two corpora include topics such as the pros and cons of animal testing and part-time jobs for university students, which require the writers to make their arguments on controversial topics. I selected ICNALE and LOCNESS for the following reasons.

Most importantly, the two corpora are compatible with learners’ data. This study attempts to rank the target bundles by combining L1 and L2 corpora. Specifically, this study compares Japanese learners’ use of lexical bundles with that of L1 English speakers under the same condition. Using this method, the difficulty affecting the prioritization of the bundles should emerge. As Granger (2017) argued, an L1 corpus alone may not be sufficient to guide the learning order of target bundles. For example, when a list presents the bundle *at the same time*, the difficulty for Japanese learners is not shown. If presented within a learner corpus, however, this would differ. If learners use the bundle with a frequency close to that of L1 English speakers, their use would be native-like, and they increase their proficiency in using the bundle. By contrast, if learners use the bundle never or fewer times than L1 English speakers, the implication would be that the bundle is difficult to acquire and requires further attention. Granger (2017) encouraged this use of a learner corpus to identify/prioritize target vocabulary items for learners with diverse L1s.

Additionally, ICNALE and LOCNESS contain what can be regarded as “target language use,” such as A-level essays in *alevels* (file) in LOCNESS and those written by instructors or professors of English (ICNALE). Granger’s (2017) study also seems to support this. She found that frequent bundles in LOCNESS and those in professional academic writing in the British National Corpus (BNC) share a higher degree of similarity. The similarity shows that L1 English speakers in LOCNESS use model-like academic English. It also suggests that L1 English speakers use multi-word expressions that exhibit the *idiom principle* (Sinclair, 1991). The principle posits that “a language user has available to him or her many of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments” (p.110). This implies that L1 English speakers use similar multi-word expressions when they express their ideas in argumentative writing (essays and research papers).

For these reasons, I considered the essays written by L1 English speakers in ICNALE and LOCNESS to be appropriate for this study, which aimed to identify target bundles for argumentative essay

writing. During the essay compilation, I excluded the essays in LOCNESS that were not argumentative, such as literary and exam essays in the file “USMIXED.” In so doing, the study secured the representativeness of the corpora. That is, all of the texts are argumentative writing.

Table 1 presents the total number of argumentative essay subcorpora and the words analyzed in this study. The files PTJ and SMK are in ICNALE, and the remaining files are in LOCNESS.

Table 1. L1 Argumentative Essay Corpus Composed for the Study

Subcorpora	No. of files	No. of words
PTJ	1	45,415
SMK	1	45,198
alevels	9	60,209
BRSUR3	1	19,019
USARG	1	149,574
USMIXED	1	9,296
Total	14	328,711

3.2. Japanese Learner Corpus

The purpose of this study was to explore the acquisition of “target bundles” by Japanese university students. For this purpose, I used the Japanese university students (JPN) corpora in ICNALE.

As mentioned in 3.1., the essays by Japanese learners in ICNALE are compatible with the L1 English speakers in the same corpus. Learners and L1 English speakers wrote their essays under the same conditions (time and topics). Moreover, ICNALE contains detailed information on each Japanese learner’s proficiency level. It ranks the learners by using the Common European Framework of Reference for Languages (CEFR). The proficiency levels from lowest to highest are A2, B1_1, B1_2, and B2, with A2 being the lowest and B2 being the highest. This study employed the developmental approach (Chen & Baker, 2013) which attempts to explore learners’ progress on language use from a developmental perspective. I had to collect each learner’s proficiency data. Because the number of essays by learners at B2 was less than 10,000 and deemed too small, I excluded the data and focused on A2 and B1 learners. For a more simplified analysis, B1_1 and B1_2 were combined and labeled “B1” in this study. Table 2 presents the data of the Japanese learner corpus.

Although the Japanese learner data in the International Corpus of Learner English (Granger et al., 2020) were compatible with those in LOCNESS (the other L1 subcorpora), I did not analyze the learners’ data. There were 10 essay topics, and combining their data with those of learners in ICNALE was difficult. ICNALE limits the topics to two (smoking and part-time job); thus, topic-related words (e.g., *part-time job*) in ICNALE strongly affect overall frequency information. For the same reason, LOCNESS was used only to extract target bundles, not for comparisons between Japanese learners and L1 English speakers.

Table 2. Japanese Learner Corpus

Subcorpora	No. of files	No. of words
A2	1	68,528
B1	1	10,1981
Total	2	170,509

3.3. Operationalization of Lexical Bundles

This study defined lexical bundles as frequently recurring clusters comprising four-words (e.g., *on the other hand, in the case of*). To distinguish between other multi-word expressions (e.g., *collocations* and *idioms*), scholars have established three statistical criteria for multi-word units to qualify as lexical bundles. A detailed account of these criteria and how this study adopted them is as follows:

Frequency: lexical bundles are blocks of words that occur at least 20-40 times per million or more in different texts (Biber & Barbieri, 2007; Hyland, 2008; Liu & Chen, 2020; Omidian et al., 2021). This shows that they appear not by chance but are sufficiently frequent to be considered linguistic phenomena. For frequency, this study used a lenient cut-off: 20 occurrences per million. This value was used because this study used a significantly smaller corpus (approximately 330,000 words) than that of two similar studies, which analyzed approximately 3 million words (Hyland, 2008) and 1 million words (Omidian et al., 2021). By setting the minimum number of frequency cut-off, this study aimed to identify lexical bundles that frequently occur even in smaller size of L1 corpus in the study.

Range: range refers to how widely lexical bundles are found in different texts. Studies have defined that lexical bundles occur in five or more texts (corpora divided from the main corpus) (Biber et al., 2004; Bychkovska & Lee, 2017; Cortes, 2004). This study followed the studies that set the minimum range of five. As aforementioned, the small-sized L1 argumentative essay corpus in this study contained approximately 330,000 words. Thus, securing the representativeness of the corpus would be necessary. Are the bundles commonly used in argumentative essay writing? As Nation (2016) argues, range is one of the most important criteria in selecting target vocabulary items. LOCNESS is useful in this regard. The number of essay subcorpora from LOCNESS collected in this study was 12. These include various topics such as the pros and cons of boxing and issues regarding British political system. Combining LOCNESS (12 subcorpora) with ICNALE (2 subcorpora), the argumentative essay corpus in this study has 14 subcorpora. I extracted the lexical bundles that occurred in five different texts in the 14 subcorpora. In light of the definition of lexical bundles, this study presents sufficient subcorpora and range criterion. This is because five occurrences in 14 subcorpora mean the bundle is found in about the half of the subcorpora (seven out of 14), indicating the wide use of the bundle.

Word length: in addition to frequency and range, I set the length of four to collect lexical bundles from the corpus. Generally, four-word bundles provide sufficient data for linguistic analysis (Omidian et al., 2021) because they have various types of fixed expressions (e.g., *at the same time, when it comes to*). This four-word criterion was adopted in several influential studies on lexical bundles (Byrd & Coxhead, 2010; Cortes, 2004; Hyland, 2008). As for this study, four-word bundles are more reliable than three- or five-word bundles to capture important noun-based bundles such as *on the part of* and *in the form of*.

The bundles that fulfilled the aforementioned three criteria of frequency, range, and length were further filtered to confirm whether they occurred in at least one of the ICNALE subcorpora (i.e., PTJ or SMK). The purpose was to use L1 English speakers' data in ICNALE as a criterion for frequency information of target bundles when compared with learners.

The concordance software AntConc performed all of the procedures described. This resulted in 78 target bundles. Figure 1 shows the extraction process of target bundles in this study.

ENS means English native (L1) speakers in ICNALE.

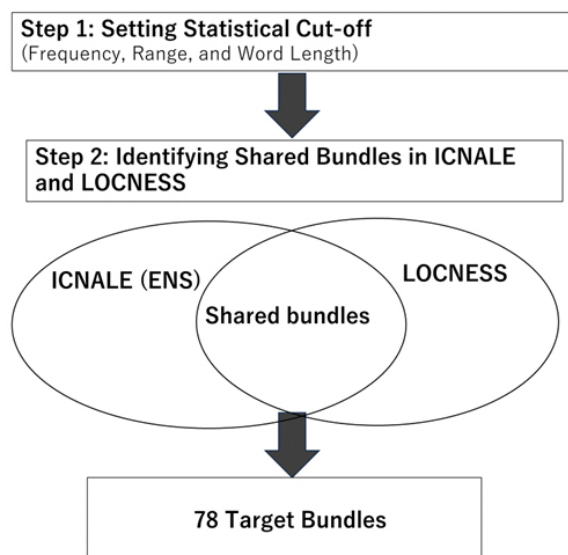


Figure 1. Extraction Process of Target Bundles

3.4. Functional Categorization

RQ 1 regarded how target bundles can be functionally classified and how Japanese learners use them. To answer RQ1, functional categories were manually annotated to the 78 target bundles. This study referred to functional taxonomies by Biber et al. (2004), Chen and Baker (2016), and Simpson-Vlach and Ellis (2010). The reasons for selecting the above three taxonomies are as follows:

Reliability as the fundamental taxonomy: Biber et al.'s (2004) taxonomy laid the foundation for Chen and Baker (2016) and Simpson-Vlach and Ellis (2010). Biber et al. (2004) identified three major categories: referential expressions, stance bundles, and discourse organizers. Referential expressions specify any entity that is after or is included in the bundles. Referential expressions can express both tangible and intangible objectives (e.g., *many of people* [tangible], *a lot of time* [intangible]). Stance bundles show the writer/speaker's thoughts or positions on the topic. Subjective bundles such as *I think it is* and *it is necessary to* are. Discourse organizers function to facilitate the discourse. For instance, *as well as the* and *on the other hand* serve to add new information and to transition to a different side, respectively.

More precise categorization: Simpson-Vlach and Ellis (2010) modified Biber et al.'s (2004) taxonomy, increasing the detail. In stance bundles, for example, Simpson-Vlach and Ellis (2010) added evaluation bundles (e.g., *it is clear/obvious that*). This category was not in Biber et al.'s (2004) epistemic category in stance bundles. Additionally, the Simpson-Vlach and Ellis (2010) categorization is more precise than that of Chen and Baker (2016), who included *it is obvious that* and *I think it is* in the same category (i.e., epistemic).

Genre similarity: regarding the similarity with this study, Chen and Baker (2016) analyzed lexical bundles in argumentative essays by Chinese learners of English. This type of writing genre is the same as that in this study, which focuses on argumentative essay writing. Notably, Chen and Baker (2016) discovered the same bundles found in this study (e.g., *all over the world, there are so many*). These bundles are typical in argumentative essay writing.

All the major categories have subcategories. For instance, referential bundles can be divided into quantifying (e.g., *a lot of*) and time and location (e.g., *at the end of*).

3.5. Rating Semantic Transparency

RQ2 regarded investigating the applicability of the FTF (Martinez, 2013) to rank target bundles in

order of priority. Because target bundles already have frequency information (number of uses in ICNALE L1 English speakers), the next step was to rate semantic transparency. The following three bullet points are examples of the semantic transparency rate of the target bundles in this study. The higher the items in terms of being transparent in meaning, the more points they receive.

- *Literal* (30 points): bundles have literal meaning (e.g., *as a result of, I do not think*).
- *Polysemous* (20 points): meanings of bundles are context-dependent (polysemous; e.g., *at the same time, there is no way*). For instance, the bundle *there is no way* has literal meaning in the sentence *there is no way to solve the problem*. By contrast, *there is no way he makes such a mistake* shows the possibility or the writer/speaker's surprise. This study considered Sykes's (2017) discussion about the similar polysemous bundle *if you look at*. As aforementioned, the bundle signals different meanings in *if you look at page fifty-five* (p. 50) and *if you look at culture as a system of values* (p.51). This study considered these polysemous bundles to be less semantically transparent than literal bundles that carry only one meaning.
- *Idiomatic* (10 points): meanings of bundles are difficult to infer from each word (e.g., *it comes down to, on the other hand*). Unlike *literal* and *polysemous* bundles, which retain literal meaning, these expressions are considered to lose their literal meanings and, thus, have the lowest semantic transparency rate.

This study referred to Barghamadi et al.'s (2023) rating procedure that ranked semantic transparency of multi-word expressions such as *collocations* and *idioms*. Although focusing on different categories of multi-word expressions, Barghamadi et al.'s (2023) rating method can be used to provide semantic transparency with numbers for statistical analysis.

4. Results

4.1. Functional Categorization of Target Bundles

Table 3 presents functional categories, semantic transparency, and frequency (times used in ICNALE per million words) of the 78 target bundles.

Table 3. Functional Categories, Transparency, and Frequency of Target Bundles

Function	Transparency	Freq per million		
		A2	B1	ENS
Referential				
1. Specification of attributes				
a) Intangible framing attributes				
in the case of	30	58	39	77
in the form of	30	0	0	22
on the part of	30	0	0	22
and the fact that	30	0	0	44
is the fact that	30	15	10	22
b) Quantifying				
the rest of the	30	0	0	99
for a number of	20	0	0	44
a large amount of	30	0	0	22
there are so many	30	0	0	11
a lot of money	30	379	480	66
a large number of	30	58	0	44

Function	Transparency	Freq per million		
the majority of the	30	0	0	22
in the number of	30	0	0	11
2. Identification & focus				
is one of the	30	131	186	66
one of the most	30	73	29	66
that there is no	30	0	0	33
that it is a	30	58	69	132
that it is not	30	102	88	66
that there is a	30	15	0	44
there will be a	30	0	10	33
there is no way	20	0	10	33
that it will be	30	0	10	22
that there would be	30	0	0	22
there would be a	30	0	0	11
one of the main	30	0	0	11
that it is the	30	15	20	11
there has been a	30	0	0	11
3. Time & location				
at the same time	20	88	108	166
for a long time	30	88	39	110
in the United States	30	15	0	99
all over the world	30	102	49	55
the end of the	10	0	0	44
in the world and	30	15	0	88
at the end of	20	0	10	55
in the long term	30	0	0	11
Stance				
1. Hedges				
it would be a	30	0	0	121
and it would be	30	0	0	44
to a certain extent	30	0	0	11
would not be a	30	15	0	11
2. Epistemic				
believe that it is	30	0	29	155
I do not think	30	88	10	143
I think that the	30	233	216	132
but I think that	30	219	265	121
I believe that it	30	0	20	110
and I believe that	30	0	0	77
I believe that the	30	0	10	44
I am sure that	30	29	39	44
think that it is	30	890	686	33
I feel that the	30	0	20	33
3. Obligation				
that it should be	30	29	29	66
they would have to	30	0	0	33
it should not be	30	0	10	22
will have to be	30	15	0	22

Function	Transparency	Freq per million		
there is no need	30	0	0	22
4. Ability & possibility				
it is possible to	30	29	10	22
have the right to	30	29	108	221
should be able to	30	0	39	166
to be able to	30	58	0	155
not be able to	30	29	29	33
would be able to	30	0	29	33
will be able to	30	263	284	22
5. Evaluation				
it is important to	30	248	294	110
it is difficult to	30	131	167	66
it is hard to	30	58	39	33
it is true that	30	146	196	11
Discourse				
1. Topic introduction & focus				
I would like to	30	117	59	66
when it comes to	10	15	10	155
that they do not	30	0	20	77
it is up to	20	0	10	55
the only way to	30	0	10	44
would like to see	30	0	0	11
2. Topic elaboration				
a) Clarification				
on the other hand	10	131	137	132
as well as the	20	15	0	44
b) Cause & effect				
as a result of	30	0	0	166
whether or not the	30	0	0	44
there is no reason	30	15	0	44
it comes down to	10	0	0	33
because it is a	30	0	10	22

Figure 2 shows the functional categories of the target bundles. The total number of target bundles used by L1 English speakers (ENS) was 78. A2 learners used 37 of the 78 target bundles. B1 learners used 43 of the 78 target bundles.

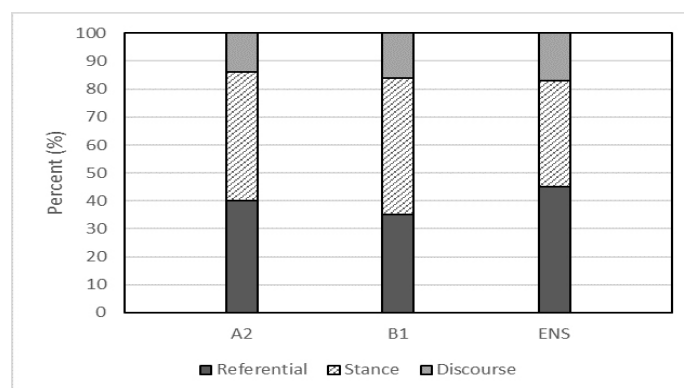


Figure 2. Functional Categorization of Target Bundles

In Figure 2, L1 English speakers (ENS) use referential bundles the most. The use of bundles amounts to 35 of the 78 bundles (45%), larger than that of A2 and B1 learners. The result supports those the literature (Juknevičienė, 2009; Nam, 2017) that learners tend to employ fewer referential bundles than L1 English speakers. The result also conforms to the hypothesis in this study that referential bundles pose a challenge for L2 learners (Japanese). This implies that target bundles used by L1 English speakers are more typical of academic writing that contains a range of referential bundles (Biber et al., 2004; Chen & Baker, 2010; Simpson-Vlach & Ellis, 2010). By contrast, the ratio of stance bundles used by L1 English speakers is smaller than that of A2 and B1 learners, whose use of the bundles was nearly 50% of the total.

Regarding the development of bundle use from A2 to B1 Japanese learners, slight increases were observed in stance and discourse bundles. When A2 learners reached B1, the use of stance bundles increased from 46% to 49%. Discourse bundles had a similar development pattern: their use increased from 14% to 16%. On the other hand, the ratio of referential bundle use declined from 41% to 35% as learners reached B1 from A2. This shows a slow development of referential bundle acquisition, which has also been found in L2 English learners (Chen & Baker, 2010; Nam, 2017; Juknevičienė, 2009; Staples et al., 2013). Details of each bundle category are presented in the following sections.

4.2. Referential Bundles

Figure 3 shows subcategories of referential bundles found in A2 and B1 learner corpora.

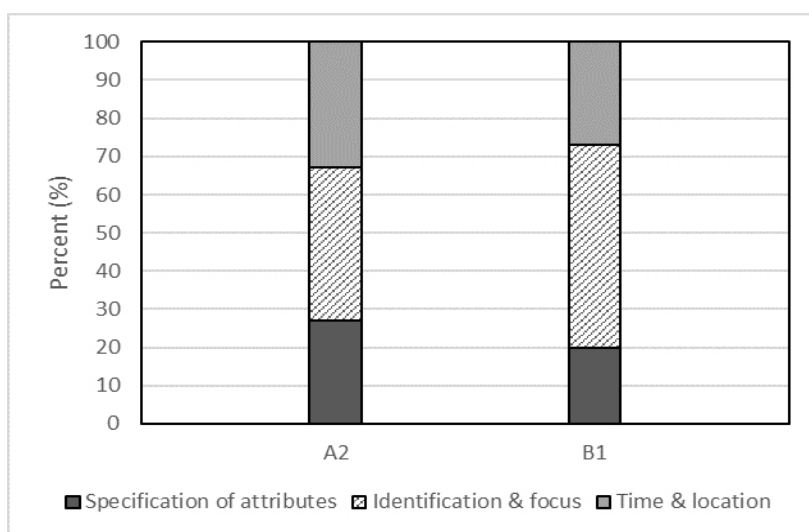


Figure 3. Subcategories of Referential Bundles

Although B1 learners are more proficient than A2 learners, only identification and focus bundles increased. Learners favored certain types of bundles despite their proficiency levels. In time & location bundles, for example, three of four bundles used by B1 learners (i.e., *at the same time*, *for a long time*, *all over the world*) had already been acquired by A2 learners. This time and location bundle dependency is universally applicable to Japanese and Chinese learners of English. The Chinese learners of English in Chen and Baker (2010) tended to use *all over the world* across proficiency levels. Similarly, the Chinese learners of English in Chen and Baker (2016) used *at the same time* and *for a long time* across three proficiency levels (low, middle, and high). This partly conforms to A2 and B1 learners' use of time and location bundles in this study.

Among the relatively limited use of referential bundles, B1 learners have better command of identification & focus bundles than A2 learners. The use of the bundles increased from 40% (A2) to 53% (B1). Locating expressions such as *there will be a* and *there is no way* seem to be first attested at B1.

4.3. Stance Bundles

A developmental feature in Figure 4 is an increase in epistemic bundles from A2 to B1.

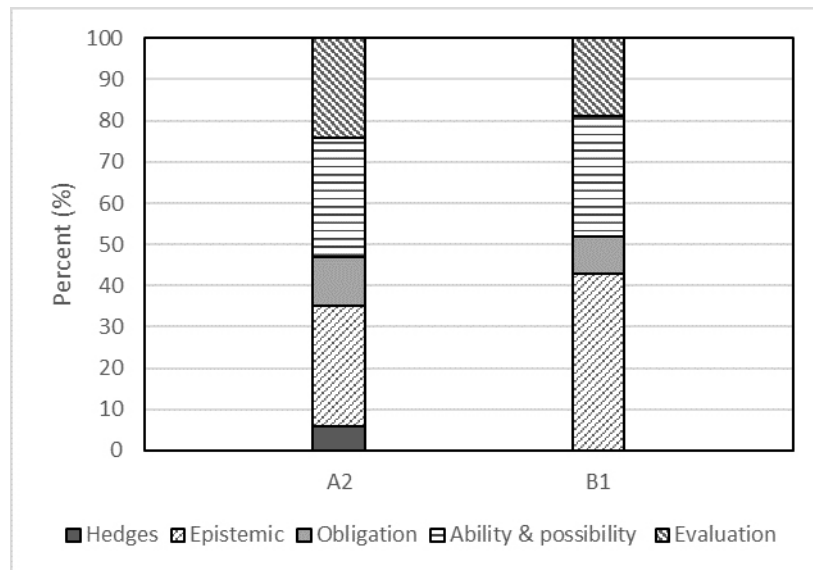


Figure 4. Subcategories of Stance Bundles

While almost all A2 learners' use of epistemic bundles is confined to *think* constructions (e.g., *I think that the, but I think that*), B1 learners acquire more verb-based epistemic bundles such as *believe that it is, I feel that the* than A2 learners. This result confirms that of Nam and Park (2020) that higher-level learners start to use more verb-based bundles.

Figure 4 shows the sustained use of ability & possibility bundles (e.g., *will be able to, it is possible to*) across proficiency levels. Syntactically similar bundles in the evaluation category (e.g., *it is necessary/difficult to*) are also favored by A2 and B1 learners. They can use all four target evaluation bundles in Table 3.

4.4. Discourse Bundles

Figure 5 represents B1 learners' significant development in topic introduction & focus bundles.

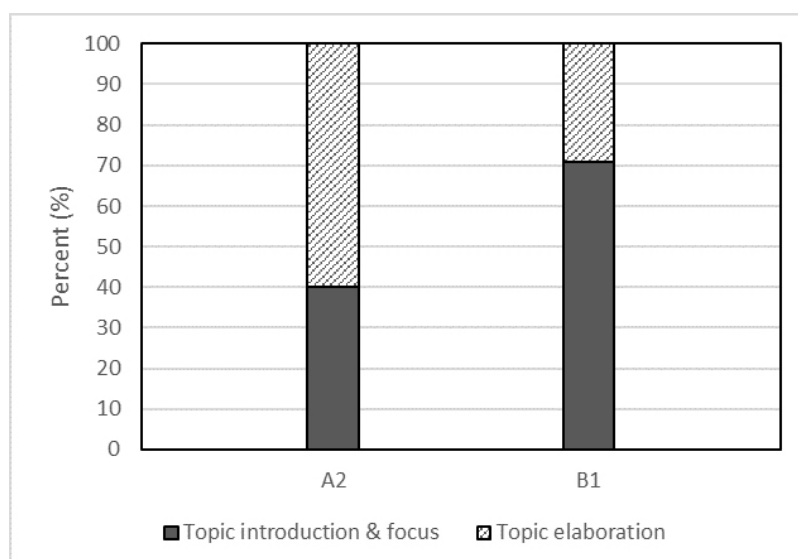


Figure 5. Subcategories of Discourse Bundles

Besides the bundles (e.g., *I would like to*, *when it comes to*) used at A2, B1 learners start to use more various and precise bundles (e.g., *it is up to*, *the only way to*). Topic elaboration bundles, by contrast, display no such development. Learners use almost no cause-and-effect bundles (e.g., *as a result of*) in topic elaboration bundles.

4.5. Application of FTF

RQ2 involved investigating the applicability of Martinez's (2013) FTF. The framework yields teaching priorities for multi-word expressions based on frequency and semantic transparency. I performed correspondence analysis by using the web app on langtest.jp to place target bundles in order of priority. Lantest.jp is a multifunctional application website that can perform a range of statistical analyses, such as cluster analysis and principal component analysis. Figure 6 shows the frequency and transparency distribution of the target bundles.

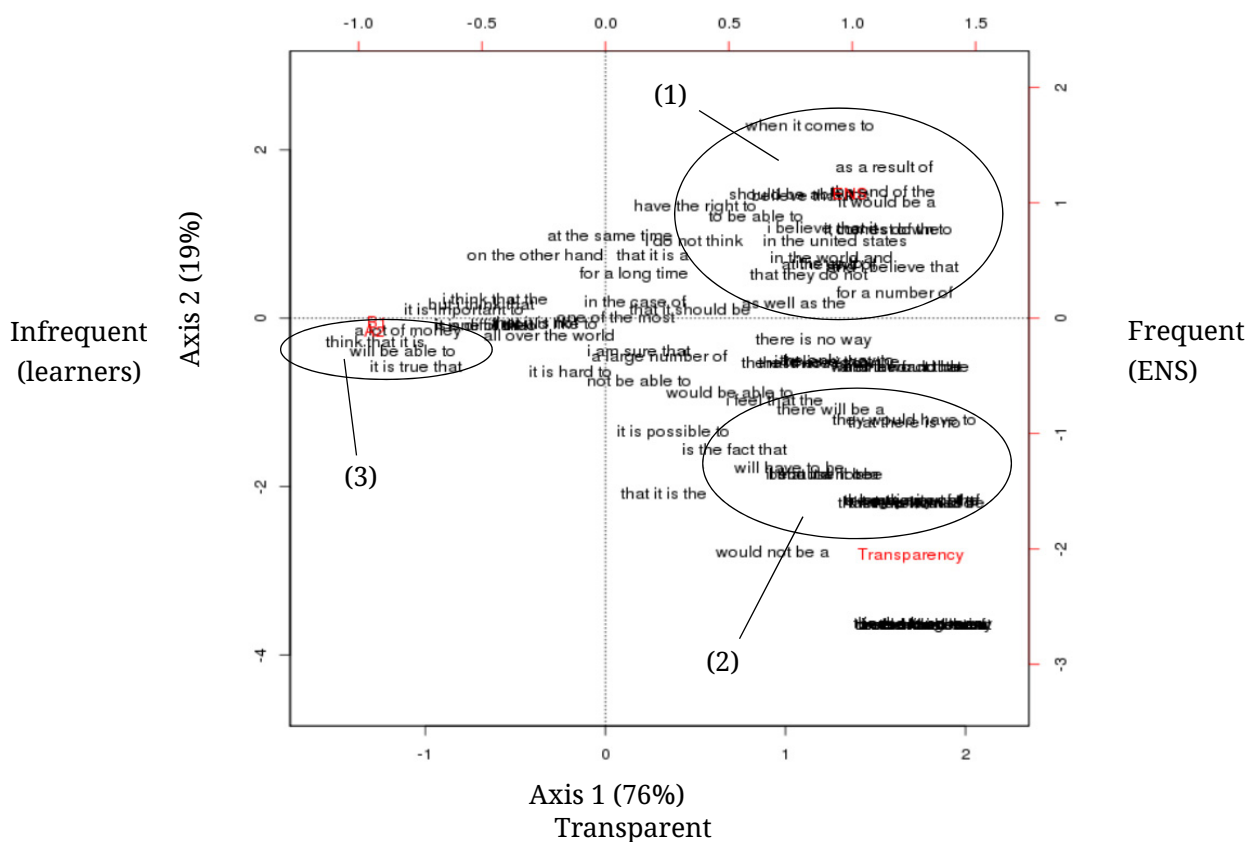


Figure 6. Distribution of Target Bundles

In Figure 6, axis 1 shows the frequency of target bundles. Axis 1 places A2 and B1 (learners) on the left, close to the horizontal line, separating them from L1 English speakers (ENS positioned at upper right) by frequency of target bundle use. This indicates that frequently used bundles greatly differ between learners and L1 English speakers. For example, the bundles *a lot of money* and *think that it is* are near A2 and B1. By contrast, bundles such as *it would be a* are just below ENS (upper right), indicating their frequent use by L1 English speakers.

Axis 2 displays the degree of semantic transparency. The higher the bundles are placed, the less semantically transparent they are. One obvious example is *when it comes to*. Its rank of semantic transparency is the lowest (10 points) and is positioned above (frequently used by) L1 English speakers.

The FTF (Martinez, 2013) showed that one effective order of teaching multi-word expressions is (1) frequent opaque→(2) frequent transparent→(3) infrequent opaque→(4) infrequent transparent

expressions. That is, learners should focus on frequent but difficult to understand (opaque) expressions first because they are necessary for receptive and productive use. By contrast, less priority would be given to infrequent and transparent expressions because they are used less in target language use (often L1 English speakers) but are easy to understand if learners use them.

As shown in Figure 6, a clear example of a lexical bundle that fits into (1) frequent opaque in the FTF is *when it comes to*, one of the discourse bundles. The bundle was located between 0.5 and 1.0 on axis 1 and nearly 2 on axis 2, indicating that the bundle was frequent and less semantically transparent. Other (1) frequent opaque bundles were *it comes down to* and *as well as the* (Figure 6). *It comes down to* was between 1.0 and 1.5 on axis 1 and close to 1 on axis 2. Although less semantically opaque than *when it comes to* and *it comes down to* (10 points in transparency), the bundle *as well as the* (20 points) was above the horizontal line; thus, it was also more semantically opaque than the bundles below the line and was more frequently used by L1 English speakers. Thus, including *as well as the* in (1) frequent opaque bundles would also be effective. Another notable bundle in (1) frequent opaque to consider is *the end of the*. Its uses by L1 English speakers were all idiomatic. They are part of an idiomatic expression *at the end of the day* (eventually, after all) and received 10 points in transparency. These bundles were located above the horizontal (opaque) line and to the right of the vertical line (frequent). This conforms to the teaching order of FTF, which placed frequent opaque items as the first priority.

The second most important bundles to teach were the (2) frequent transparent bundles positioned in the lower right of Figure 6. These bundles were frequent (inclined toward right) and transparent (under the horizontal line). For example, *there will be a* is a semantically transparent identification & focus bundle, but learners used it less often. This bundle only had literal meaning; thus, learners would have less burden when studying it. Frequent transparent bundles are functionally important as well. They include the specification of attribute bundles that learners do not seem to acquire. These are *the majority of the*, *in the form of*, and *on the part of*. These noun-based bundles are typical of academic writing (Granger, 2017); thus, they are also important bundles.

Regarding (3) infrequent opaque bundles, it is difficult to locate such bundles in Figure 6. Unlike (1) frequent opaque and (2) frequency transparent bundles, positioned at the upper right or lower right of axis 1, the upper left has few word clusters. Although opaque bundles (e.g., *on the other hand*) are displayed around the upper center of Figure 6, they were not infrequently used. They were frequently used by learners (A2 and B1) and L1 English speakers. Their use was 131 times by A2, 137 times by B1, and 132 times by L1 English speakers per million. Thus, their tendency was neither toward the left nor the right in frequency. Learners had a good command of the opaque bundles positioned at the center of Figure 6, at least in frequency. Therefore, this study did not target (3) infrequent opaque bundles; instead, it included infrequent transparent bundles as the third bundles for discussion. According to Martinez (2013), this type of modification is acceptable. “[i]t is entirely possible that all of the multi-word expressions in a given text fit into just one or two of the FTF quadrants.” (p. 192).

(3) the infrequent transparent bundles in Figure 6 are *a lot of money*, *think that it is*, *will be able to*, and *it is true that*. They are less frequently used by L1 English speakers and too frequently used by both A2 and B1 learners. Additionally, they are semantically transparent (literal in meaning). Learners’ preferences of the bundles (frequency and semantic transparency) have strong correlation with each other ($r = .95$). Specifically, *think that it is* and *a lot of money* would be caused by L1 (Japanese). As discussed in 4.3. (stance bundles), *think* construction (～と思う) is commonly observed in Japanese. Regarding *a lot of money*, learners were probably affected by Japanese 多くの (a lot of) construction. Martinez (2013) cautioned that learners’ L1 equivalence affects the priority of teaching order. In the case of Japanese learners, the aforementioned bundles should be the last target for learning because learners have already acquired these bundles and are likely to retain them with the help of L1 translation.

5. Discussion

RQ1 of the study sought to functionally categorize target lexical bundles and what items learners should focus on. The results revealed that while learners acquire more stance and discourse bundles as their proficiency levels advance, the same is not true for referential bundles. Referential bundles should be the priority in terms of their functions.

Among the use of the three subcategories in referential bundles (Specification of attributes, identification & focus, and time & location), only identification & focus displayed the increase. This slow development might be a cause for the decline in the ratio of total referential bundle uses in A2 (41%) and B1 (35%) in Figure 2. Continued dependency on time & location bundles also restricts the total uses of referential bundles. One possible explanation for the dependency is learners' L1 influence. One obvious example is *for a long time*. It has a direct translation to the Japanese phrase 長い時間 (*long* = 長い: *nagai*, *time* = 時間: *jikan*). Thus, Japanese learners might be comfortable in using the bundle. As Yamashita and Jiang's (2010) study demonstrated, learners' L1 promotes the acquisition in this case. Although the tendency to use time and location bundles seems to have led to the limited variations of referential bundles across proficiency levels, the use itself should be accepted. This is because all the bundles analyzed in this study are "target" bundles that fulfilled statistical criteria in the L1 English corpus and seem to be worth learning. This positive L1 influence can also be found in specification of attributes bundles as well. In contrast to *in the form of* and *on the part of*, learners used more *in the case of* bundles. Since Japanese has the word ケース (direct translation of the English *case*), this equivalence could facilitate the use. Thus, while using such bundles, Japanese learners should target other bundles that deserve more attention.

Specification of attributes can be the targets for more varied use of referential bundles. Specifically, noun-based bundles (e.g., *in the form of*, *on the part of*) are worth noting because they are typical of academic writing but are also underused by not only Japanese learners but also learners of different L1s (Granger, 2017; Nam & Park, 2020).

As the better use of identification & focus bundles in referential bundles is a feature of B1 learners, teaching these bundles to A2 learners might be effective in their proceeding to B1.

Semantic aspect of referential bundles should also be considered for the inclusion of target bundles. For instance, *the end of the* in time and location bundles is a part of a semantically opaque idiomatic expression *at the end of the day* (eventually), and learners seem to be unfamiliar with it. This complicated nature of referential bundles may be one of the reasons for the slow development. Although most of the target referential bundles in this study are semantically transparent, opaque ones such as the above *at the end of* should be taken into account.

Regarding stance bundles, a prominent feature is that B1 learners acquire more epistemic bundles than A2 learners. As discussed in referential bundles, a positive L1 influence (Japanese) also emerges in epistemic bundles. Japanese epistemic constructions have a range of verb-based constructions similar to that-clause constructions in English. ～だと思ふ/感じる/信じる can be translated into *I think/feel/believe* that, respectively. These similarities can promote the acquisition of verb-based target epistemic bundles. By contrast, studies have reported the overuse of these verb-based bundles, such as *think* (Ishikawa, 2011; Yong et al., 2010). These studies reveal that the overuse of verb-based bundles makes learners sound less native-like. Thus, this study's position is that the use of verb-based bundles should be encouraged as long as they are used not too frequently.

Lastly, B1 learners start to use more varied topic introduction & focus bundles in discourse bundles than A2 learners. However, topic elaboration bundles observe no such development. This implies the higher degree of difficulty for acquisition. As in the case of referential bundles, another crucial feature of discourse bundles is that they are often semantically opaque. Over one-third of discourse target bundles are not transparent. For instance, *when it comes to* and *it comes down to* are idiomatic, and their meanings are difficult to infer from each word. This necessitates explicit instructions for A2 and

B1 learners.

The results for RQ1 support the hypotheses in the study that referential and semantically opaque bundles are difficult for learners to acquire.

RQ2 of the study investigated the applicability of the FTF framework. Using correspondence analysis, the study set the teaching order of the target bundles to (1) frequent opaque→(2) frequent transparent→(3) infrequent transparent.

The summary of the findings for RQ1 and 2 and suggestions based on them are as follows:

RQ1: Learners' acquisition of functionally categorized target bundles

- Learners acquire more stance and discourse bundles as their proficiency level advances.
- Referential bundles exhibit a slow development for acquisition.
- Referential bundles would be the target functional category among the three.

RQ2: Teaching order of target bundles per the FTF

1. Frequent and opaque discourse bundles (e.g., *when it comes to*)
2. Frequent transparent referential bundles (e.g., *there will be a, in the form of*)
3. Infrequent transparent (L1-supported) bundles (e.g., *a lot of money, think that it is*)

Figure 7 is a teaching order suggestion for Japanese learners based on the results for RQ2.

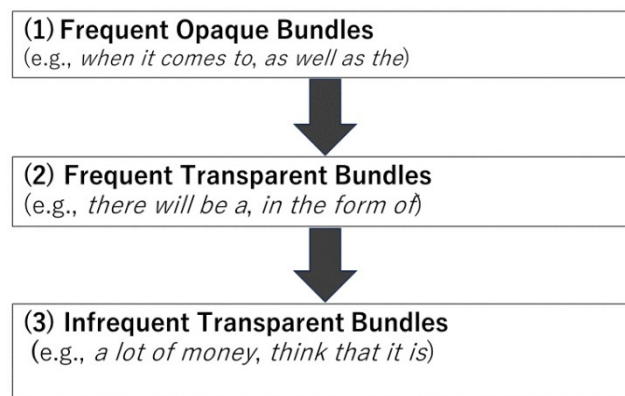


Figure 7. Teaching Order Suggestion for Target Bundles

(1) frequent opaque bundles are opaque and frequently used by L1 English speakers but not by learners. Noticeable examples of these bundles are *when it comes to, as well as the*. They are also one of the discourse bundles (topic introduction & focus, topic elaboration) that learners are not familiar with.

(2) frequent transparent bundles are transparent bundles that learners use more sparingly than L1 English speakers do. Considering the functional importance of referential bundles in academic writing, teaching them would be effective, taking advantage of their greater semantic transparency than opaque bundles in discourse bundles.

(3) infrequent transparent bundles are transparent and favored by learners. Some of them can be easily accessible to learners because of their L1. These bundles seem to be easier to acquire than the aforementioned two types of bundles. Thus, this type received the least priority in Figure 7.

Caution should be used regarding the aforementioned suggestions for teaching argumentative lexical bundles. The suggestions would vary according to students and needs. For example, the learning burden of (1) frequent opaque bundles would differ for learners with diverse learning backgrounds; as in the case of *on the other hand*, it is frequent and is a discourse opaque (idiomatic) bundle, but Japanese learners use it with a frequency similar to that of L1 English speakers. Thus, *on the other hand* is not in (1) frequent (limited to the use by L1 English speakers) opaque bundles in

Figure 6. Focusing on learners from different backgrounds (e.g., L1 or types of input) may yield different results. The same can be true for (3) infrequent transparent bundles. For Japanese learners, *think* construction in the bundles seems to have a strong influence from Japanese, which may differ for learners with other L1s.

Another issue is related to possible discrepancies in the framework. Because most lexical bundles are semantically transparent (e.g., *as a result of*), neatly categorizing all the bundles into the FTF was impossible. For instance, the bundle *as a result of* (transparent) was closely located to *when it comes to* (1: frequent opaque) (Figure 6) because of their similarity in frequency. In this case, distinguishing between the two is necessary when considering the inclusion of the bundles.

Summarizing the results for RQ1 and 2, this study proposes that, for the inclusion of target bundles of argumentative essay writing for Japanese learners;

- referential bundles should be the target bundle category in terms of function, and
- discourse bundles might be the first priority in terms of semantic opacity and frequency.

The above two proposals could be incorporated into EAP vocabulary teaching practice and materials development for Japanese learners.

In vocabulary teaching, a practitioner can prioritize referential and discourse bundles when teaching essay writing to A2 and B1 learners as both groups may not be familiar with the bundles despite their significance in terms of functions and semantic transparency. Meanwhile, particular emphasis could be put on stance and discourse bundles if a teacher helps A2 learners proceed to the next level, B1. As discovered in this study, more varied uses of stance and discourse bundles are characteristics of B1 learners. Harrison and Barker also (2015) suggest this step-by-step vocabulary teaching to learners at different proficiency levels.

For EAP teaching materials development, developers can refer to this study when selecting multi-word expressions in CEFR-based course materials such as textbooks.

6. Conclusions

This study attempted to identify the target lexical bundles for argumentative essay writing, an important area in EAP, and discovered 78 target bundles. Specifically, this study provides insights into two features of lexical bundles: discourse functions and semantic transparency. After employing L1 and L2 corpora in ICNALE, this study proposes that, among the target bundles, Japanese learners focus on referential bundles in terms of their functional characteristics. From the view of semantic transparency, teaching priority would be frequent but unfamiliar semantically opaque bundles to learners.

The study is novel in that it effectively combined L1 and L2 corpora, taking advantage of both corpora. An L1 corpus is valuable to provide target-like language use, but relying solely on it results in missed pedagogical issues such as the learnability for learners. Similarly, using the L2 corpus alone may not present target items if there is no attempt to extract them from the L1 corpus. The L1 and L2 corpora combined in this study should complement each other. Additionally, this study accepts what can be referred to as negative bundle use. For example, studies have regarded *I think* construction as being less-sophisticated or have limited use by learners. However, this study demonstrated that *think* bundles are the target bundles that L1 English speakers use widely and frequently. Thus, instead of completely avoiding the use of such bundles, this study suggests that they be ordered by priority (e.g., *think that it is* for Japanese learners).

The study has limitations. First, as acknowledged in the method section, the target bundles are from a small L1 corpus (approximately 330,000 words). The primary objective of this study was to identify target bundles by combining L1 and L2 corpora. Thus, although the number of words is not the main

concern, I must acknowledge that the corpus size was smaller than in other studies that set target vocabulary items from 1 million words or more. This necessitates further research for generalization. Second, although not limited to this study, classifying lexical bundle functions and rating semantic transparency inevitably involved subjectivity.

Despite its limitations, this study makes pedagogical and methodological contributions to the application of corpus in college-level EAP research and practice.

Acknowledgments

This study was supported by JST SPRING (grant number JPMJSP2150).

References

- Ackermann, K., & Chen, Y. H. (2013). Developing the academic collocation list (ACL): A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, 12(4), 235-247.
- Appel, R., & Murray, L. (2020). L1 differences in L2 English academic writing: A lexical bundles analysis. *Journal of English for Academic Purposes*, 46, 1-15.
- Barghamadi, M., Rogers, J., Arciuli, J., & Müller, A. (2023). The use of semantic transparency and L1-L2 congruency as multi-word units selection criteria. *Studies in English Language and Education*, 10(2), 723-740.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26, 263-286.
- Biber, D., S, Conrad., & Cortes, V. (2004). "If you look at ...": Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25, 371-405.
- Boers, F., & Lindstromberg, S. (2009). *Optimizing a Lexical Approach to Instructed Second Language Acquisition*. London: Palgrave Macmillan.
- Bychkovska, T., & Lee, J. J. (2017). *At the same time*: Lexical bundles in L1 and L2 university student argumentative writing. *Journal of English for Academic Purposes*, 30, 38-52.
- Byrd, P., & Coxhead, A. (2010). *On the other hand*: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, 5(5), 31-64.
- Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology*, 14(2), 30-49.
- Chen, Y. H., & Baker, P. (2016). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics*, 37(6), 849-880.
- Conklin, K., & Schmitt, S. (2008). Formulaic Sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72-89.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23, 397-423.
- Council of Europe. (2001). *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Gardner, S., Nesi, H., & Biber, D. (2018). Discipline, level, genre: Integrating situational perspectives in a new MD analysis of university student writing. *Applied Linguistics*, 40(4), 646-674.
- Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. In Granger, S. (Ed.), *Learner English on Computer*. Addison Wesley Longman: London & New York, 3-18.

- Granger, S. (2017). Academic phraseology: A key ingredient in successful L2 academic literacy. In Vatvedt Fjeld, R., Hagen, K., Henriksen, B., Johannson, S. Olsen, S., & Prentice, J. (Eds.), *Academic Language in a Nordic Setting: Linguistic and Educational Perspectives*. *Oslo Studies in Language*, 9(3), 9-27.
- Granger, S. (2019). Formulaic sequences in learner corpora: Collocations and lexical bundles. In Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (Eds.). (2019). *Understanding Formulaic Language: A second Language Acquisition Perspective* (pp. 228-245. New York, NY: Routledge.
- Granger, S., Dupont, M., Meunier, F., Naets, H. & Paquot, M. (2020). *The International Corpus of Learner English*. Version 3. Louvain-la-Neuve: Presses universitaires de Louvain.
- Harrison, J., & Barker, F. (2015). *English Profile in Practice*. Cambridge: Cambridge University Press.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.
- Ishikawa, S. (2011). Phraseology overused and underused by Japanese learners of English. In Yagi, K., Kanzaki, K., & Inoue, A. (Eds.). *Phraseology, Corpus Linguistics and Lexicography: Papers from Phraseology 2009 in Japan* (pp.82-93). Hyogo: Kwansei Gakuin University Press.
- Ishikawa, S. (2023). *The ICNALE Guide: An Introduction to a Learner Corpus Study on Asian Learners' L2 English*. London: Routledge.
- Johnson, D. (2018). Teaching English for academic purposes in New Zealand: Making sense of genre-based instruction. In Wong, L.T., & Wong, Heidi. W. L. (Eds.), *Teaching and Learning English for Academic Purposes: Current Research and Practices* (pp. 239-253). New York, NY: Nova Science Publishers.
- Juknevičienė, R. (2009). Lexical bundles in learner language: Lithuanian learners vs. native speakers. *Kalbotyra*, 61(3), 61-72.
- Khodadady, E., & Shamsaee, S. (2012). Formulaic sequences and their relationship with speaking and listening abilities. *English Language Teaching*, 5(2), 39-49.
- Laufer, B. (2021). Formulaic sequences and second language learning. In Szudarski, P., & Barclay, S. (Eds.), *Vocabulary Theory, Patterning and Teaching* (pp. 89-98). Bristol: Multilingual Matters.
- Liu, C. Y., & Chen, H. J. H. (2020). Analyzing the functions of lexical bundles in undergraduate academic lectures for pedagogical use. *English for Specific Purposes*, 58, 122-137.
- Martinez, R. (2013). A framework for the inclusion of multi-word expressions in ELT. *ELT Journal*, 67(2), 184-198.
- Nam, D. (2017). Functional distribution of lexical bundle in native and non-native students' argumentative writing. *The Journal of Asia TEFL*, 14(4), 703-716.
- Nam, D., & Park, K. (2020). Lexical bundles as criterial features in L2 academic writing: structural differences between CEFR A2 and B2 essays. *Multimedia-Assisted Language Learning*, 23(3), 68-86.
- Nesi, H., & Matheson, N., & Basturkmen, H. (2017). University literature essays in the UK, New Zealand and the USA: Implications for EAP. *New Zealand Studies in Applied Linguistics*, 23(2), 25-38.
- Omidian, T., Siyanova-Chanturia, A., & Durrant, P. (2021). Predicting parameters of variation in the use of academic multiword expressions in university student writing. In Szudarski, P., & Barclay, S. (Eds.), *Vocabulary Theory, Patterning and Teaching* (pp. 141-166). Bristol: Multilingual Matters.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In Richards, J. C., & Schmidt, R. W. (Eds.), *Language and Communication* (pp. 191-225). London: Longman.
- Sawaguchi, R., & Mizumoto, A. (2022). Exploring the use of *make* + noun collocations by Japanese EFL learners through a bilingual essay corpus. *Corpora*, 17(SI), 61-77.

- Schmitt, N. (Ed.). (2004). *Formulaic Sequences: Acquisition, Processing and Use* (pp. 269-292). Amsterdam: John Benjamins Publishing Company.
- Shamsabadi, R., Ketabi, S., & Eslami Rasekh, A. (2017). Developing Iranian EAP students' writing skill through explicit instruction of lexical bundles. *Journal of Language Teaching and Learning, 19*, 25-52.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(12), 487-512.
- Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (Eds.). (2019). *Understanding Formulaic Language: A Second Language Acquisition Perspective*. New York, NY: Routledge.
- Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes, 12*, 214-225.
- Sykes, D. (2017). An investigation of spoken lexical bundles in interactive academic contexts. [Master's thesis, Carleton University].
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wu, S. M. (2006). Creating a contrastive rhetorical stance: Investigating the strategy of problematization in students' argumentation. *REC journal, 37*(3), 329-353.
- Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly, 44*(4), 647-668.
- Yong, W., Jingli, W., & Zhou, C. (2010). The use of *I think* by Chinese EFL learners: A study revisited. *Chinese Journal of Applied Linguistics, 33*(1), 3-23.

THE AUTHOR

Ryo Sawaguchi is a PhD student in Foreign Language Education and Research at Kansai University. His research focuses on the pedagogical applications of corpus linguistics and vocabulary learning/teaching in the English for academic purposes context.

THE AUTHOR'S ADDRESS

First and Corresponding Author

Ryo Sawaguchi

PhD Student

Graduate School of Foreign Language Education and Research, Kansai University

3-3-35 Yamate-cho, Suita-shi, Osaka 564-8680, JAPAN

E-mail: rswgch@gmail.com

Received: 11 June 2024

Received in Revised Form: 3 August 2024

Accepted: 10 August 2024