# A corpus-based comparison of EFL, ESL, and ENS lexical bundles in the ICNALE corpus

## Prihantoro[1*], and Shin'ichiro Ishikawa[2]

[1]Department of Linguistics, Faculty of Humanities, Universitas Diponegoro, Jl. Dr. Antonius Suroyo, Tembalang, Kec. Tembalang, Kota Semarang, Jawa Tengah 50275, Indonesia
[2]IPHE, Graduate School of Intercultural Studies, Kobe University, 1-2-1, Tsurukabuto, Nada-ku, Kobe, Japan

**ABSTRACT**

The importance of lexical bundles (LBs) to shape academic writing has been extensively studied. However, few studies have investigated learners of English as a Foreign Language (EFL) and English as a Second Language (ESL) alongside Native Speakers (ENS) using the same protocol. To address this gap, we investigated ICNALE's (International Corpus Network of Asian Learners of English) dataset, focusing on three subsets. Building on Biber et al.'s (2004) and Hyland's (2008) LB seminal works, we examined their frequency, structure, and function. The findings show a strong positive relationship between LB use and students' proficiency levels. In the ENS subset, the use of LBs is relatively more varied and frequent. Across the three subsets, we also discovered that prepositional phrases and text-oriented LBs were most common. This shows the roles of these LBs in managing discourse and preserving coherence. This also indicates that the LB profile may serve as a descriptor of a learner's status. It calls for integrating explicit LB instruction into the teaching of writing in Asian EFL and ESL settings. The results can be used to support curriculum design (e.g., LB-related tasks in academic writing syllabi) and materials development (e.g., high-frequency LB-based glossaries or phrase banks), among other uses. These applications can help students write more effectively and cohesively, and provide direction for ongoing or future corpus-based studies.

**Keywords:** Corpus linguistics; education; EFL; ESL; lexical bundle

## INTRODUCTION

While scholars have proposed various definitions of lexical bundles (LBs) – as shown, for example, in Flowerdew and Mahlberg (2009), Libben (2022), and Hyland and Bondi (2006) – they agree that an LB is a sequence of words that frequently occur together, such as 'on the one hand', 'on the other hand', or 'it is important to', regardless of their grammatical, semantic, or idiomatic status.

The prominence of LBs is attested across languages such as Indonesian (Budiwiyanto, 2023), Korean (Kim, 2009), and Portuguese (Matte & Goulart, 2020). However, the majority of LB studies are centred on the English language, particularly in academic contexts. Language scholars argue that LBs are important not only to effectively construct phrases and sentences (Kurniawan & Haerunisa,

2023; Li et al., 2024; Yoo & Shin, 2022) but also to build discourse coherence (O'Flynn, 2022; Oktavianti & Prayogi, 2022; Zahra et al., 2021) and express stance (Fitriati & Wahyuni, 2019; Liu & Chen, 2020; Yakut & Yuvayapan, 2022). Studies on how learners of English use LBs can shed light on their language proficiency, as shown in Appel and Murray (2023), Shin (2018), and Li and Volkov (2018). In addition to proficiency, LBs offer students practical lexical sets to profile their writing to match a specific discourse community (by nativity, professionalism, genre/domain specificity). This is evident in Kurniawan and Haerunissa's study (2023), which showed differences in the LBs present in accepted and rejected manuscript submissions. It also echoes Hyland and Jiang's

*Corresponding author
Email: prihantoro@live.undip.ac.id

(2018) argument that LBs may serve to mirror users' communicative experiences.

The frequency of lexical bundles can be used to characterise spoken and/or written proficiencies, as attested in Biber (2004), Kwary et al. (2017), Ibrahim and Khalit (2020), and Yan (2022). Li and Volkov (2018) conducted an LB study on the writing section of a Canadian English Language Proficiency programme. They showed that test-takers with higher proficiency skills tend to use more frequent text/ place/ time reference LBs. Chen and Baker (2010), who compared the LBs of L1 and L2 English authors, argue that higher frequency bundles are prevalent in academic texts, indicating their significance in conveying complex ideas.

In LB studies, in addition to measuring the frequency of individual LBs, it is also common to measure their frequency by structural or functional type, as done by Qi and Pan (2020), Hyland and Jiang (2018), and Kurniawan and Haerunissa (2023). LBs can be grouped into three main structural types: PP (prepositional phrase), NP (noun phrase), and VP (verb phrase), a categorisation proposed by Biber et al. (2021). Using this, Chen and Baker (2010) managed to show the relationship between structural types and proficiency. They argue that expert writers use more NP bundles than novice writers in a university setting. Kang et al. (2024) frequently used LBs to create teaching materials for Korean students (EFL) by targeting preposition errors. This shows that these structural categorisations also have practical pedagogic implications. Hyland (2008) and Biber et al. (2004, 2021) are two seminal works on functional types of LBs.

They share that functional types can be classified into 'broad' and 'narrow' categories. For example, certain lexical bundles, such as 'in the case of' and 'a large amount of', are 'framing' and 'quantification' categories in terms of Biber et al.'s (2004, 2021) functional types. These two categories are subsumed under 'referential' LBs, which are used to reference entities, properties, and quantities. The other two broad categories are stance expressions and discourse organisers. Hyland (2008) developed another taxonomy: research, text, and participant-oriented LBs. Likewise, each broad category subsumes more specific functional types. For instance, research-oriented bundles cover location (e.g., 'at the end of') and quantification (e.g., 'one of the most'). As mentioned above, functional types may be used to profile a text or to create a text that matches the profile of a discourse community. Chenda et al. (2022) explain that stance LBs are more frequently found in the introduction to a research paper. Estaji and Montazeri (2022) argue that referential function occurs more frequently in the introduction and discussion parts.

Candarli and Jones (2019) argue that it is possible to find correspondence between Hyland's and Biber et al.'s functional types. They explain that what Hyland (2008) refers to as research, text, and participant-oriented LBs corresponds to Biber et al.'s (2004) referential, discourse organiser, and stance expressions, respectively. If we follow this argument, the two labels are interchangeable. But for consistency, we prefer to use Hyland's labels in this paper. This is primarily because Hyland's categorisation is driven by written data, which fits the written data used in this research. Both spoken and written corpora drive Biber et al.'s categorisation.

Research on LBs has been conducted either by comparing native and non-native authors, or by separately targeting certain domains/genres, or language groups. In EFL studies, for instance, Yu's (2023) study found limited use of lexical bundles among Chinese EFL learners, suggesting a proficiency gap. Yoo and Shin's (2022) study examined the LBs used by Korean college students at different proficiency levels. They found a positive correlation between the degree of academic LBs and proficiency level. Appel and Murray (2023) studied the LBs of academic writers from three EFL countries (China, Korea, and Japan). They discovered certain country traits: a tendency among Chinese writers to overuse text-oriented LBs, among Japanese writers to use participant-oriented LBs, and among Korean writers to use LBs in general (types and tokens). This suggests that proficiency and native writing models may affect the use of these countries' EFL learners' LBs.

In ESL settings, we can consider task types. Zahra et al. (2021) studied English textbooks in Pakistan and discovered that discourse organisers and referential expressions were the most frequently used functional categories. Ibrahim and Khalit (2020) studied open-ended and graphic-oriented essays in a Malaysian ESL corpus. They discovered that more varieties were used in the former. This indicates that in an ESL setting, LBs use may also vary by task type rather than proficiency, as observed earlier in the EFL setting.

Various studies specifically targeting ENS groups have been conducted. Here, we highlight O'Flynn (2022), who discovered 47 very frequent lexical bundles across disciplines. He created a tool to automatically detect LBs and incorporate those into L2 teaching materials. Unlike other LB studies, O'Flynn's corpus was large and covered a wide range of disciplines. Simpson-Vlach and Ellis (2010) studied a combination of spoken and written large corpora, such as the BNC (British English) and MICASE (American English). They offer a number of core formulas common in all academic disciplines. These studies underscore that a stable set of LBs may be present across domains.

Some LB studies have compared native and non-native LBs. To illustrate, Shin (2018) studied native and non-native English freshmen and

discovered that academic writers demonstrated similar patterns in using lexical bundles regardless of their native-English status. Fajri et al. (2020) compared L1 and L2 English academic authors. They concluded that L2 authors used fewer LBs but showed a trajectory toward native-English-like LBs. Narkprom and Poochaorensil (2022) compared dissertations written by Thai and English native authors. They argue that Thai authors overuse certain LBs, and their motives might be institutional. However, in terms of function, the two groups show a similar trend, which might stem from the same dissertation-writing convention. Guan (2022), who compared the argumentative writing of native Chinese and English authors, argues that Chinese authors match the profile of lower-proficiency authors. This shows how native and non-native authors may converge or diverge.

Overall, the review confirms that LB use, proficiency, and writing conventions are connected. However, EFL, ESL, and ENS groups have been studied separately, with different data sets and research protocols, such as only EFL (Shin, 2018; Yu, 2023), ESL (Ibrahim & Khalit, 2020; Zahra et al., 2021), or ENS (O'Flynn, 2022; Simpson-Vlach & Ellis, 2010; Tománková, 2016), as it is useful to profile the configuration of LBs in separate contexts. However, this is considered a handicap if a project aims to arrive at a simultaneous overview of the three groups (ESL, EFL, ENS). This is because the data for the aforementioned studies were not necessarily acquired under the same data-collection protocol. Also, their analytic variables may vary to some extent. While many comparative studies (Ädel & Erman, 2012; Latif et al., 2022; Salazar, 2014) offer potentially useful outcomes to characterise

native and non-native English authors' LBs, there seem to be no studies that include countries in Asia as their objects and specify the status of English as ESL and EFL in those countries, with a comparison to an ENS group. The absence of such projects leaves a research gap, which this study aims to fill. To obtain a more comprehensive overview of how LBs are used by learners in Asia (ESL and EFL) compared to ENS groups, a consistent data-collection protocol is needed. Likewise, a standardised cross-group comparative analysis between Asia vs ENs groups and ESL vs EFL groups needs to be implemented. We operationalise these in our study to fill these gaps. A global significance of this study is that our LB findings may enrich world Englishes studies, informing how English has developed cross-nationally. This serves as the background for our research question, formulated as follows: In terms of the LBs used by students from ENS and Asian background groups (ESL and EFL), how do 1) individual LBs and their 2) structural and 3) functional types differ across these groups?

## METHOD
### Corpus Data Collection and Indexing
The research procedure implemented in this project is summarised in Figure 1. Data for this research were collected from the official website of ICNALE (Ishikawa, 2023) (see Figure 2) – a learner-corpus network whose data were primarily obtained from learners of English in Asia, in both ESL (the Philippines, Singapore, Pakistan, Hong Kong) and EFL (China, Indonesia, Japan, Korea, Thailand, Taiwan) groups.

**Figure 1**
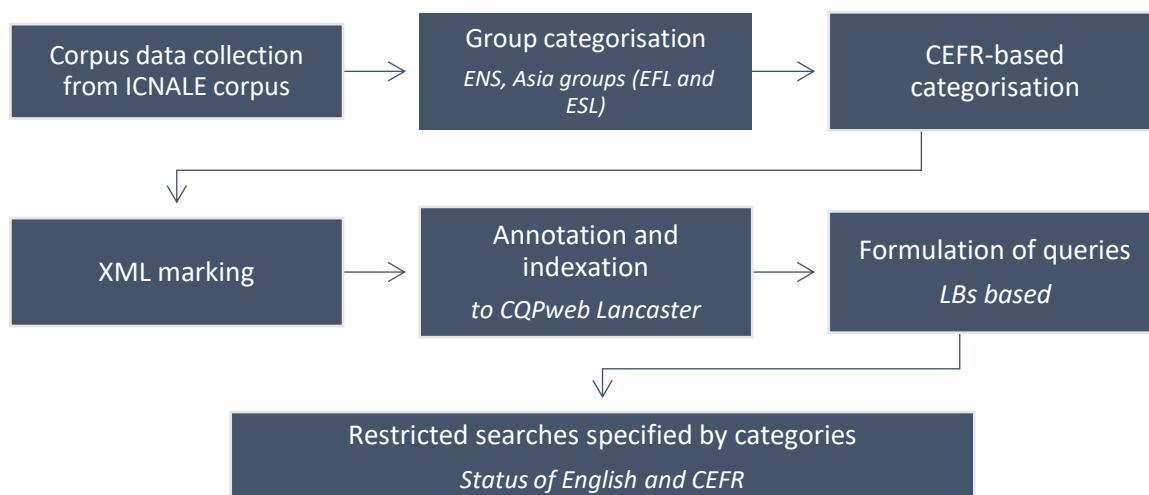*Summary of the Research Procedure Implemented in This Project*

**Figure 2**
*ICNALE Download Page*



**From existing ICNALE modules, we selected** ICNALE essays and downloaded all of them, along with their corresponding metadata (information about the respondents who submitted essays, such as country of origin, English language proficiency, and gender). Note that the corpus contains data not only from learners of English in Asia, but also from native speakers of English (ENS) in Canada, the United States, Great Britain, Australia, and New Zealand. Students were asked to write argumentative essays on two selected topics. While argumentative, the essays were not research-based or structured like response papers, theses, dissertations, or journal articles, which adhere to certain rhetorical stylesheets. The full version of the data collection protocol can be found on ICNALE's website (https://language.sakura.ne.jp/icnale/, last accessed 15 June 2024).

Data and metadata were formatted as an XML Document (as implemented in Prihantoro et al. 2025; Prihantoro & Gillings, 2025), a common document format readable, a common document format readable by popular corpus-query systems, such as LancsBox X, WordSmith, AntConc, CQPweb (Hardie, 2012, 2024), Sketch Engine (Kilgarriff et al., 2008, 2024), and English-Corpora, as surveyed by Rodrigues (2020). The first three systems are standalone applications, while the latter three are web-based.

This project not only produces an academic paper but also a corpus accessible via an open-corpus query system with a robust search function, another expected outcome that allows users to replicate or adapt the analytic procedure outlined here. Therefore, web-based applications are preferred to avoid installation problems for users. Users are only required to create an account to access the indexed corpus. Next, while all the aforementioned systems can read XML documents, the extent to which they can parse them varies. For parsing purposes, web-based applications are preferred because they typically parse XML document formats well. This boils down to three choices: CQPweb, Sketch Engine, and English Corpora.

Of these three, CQPweb (Lancaster) was preferred for data indexing for several reasons. CQPweb offers free access without creating a trial account or paying a subscription fee. Our aim is to make the corpus available free of charge, as the original data have the same licence (see Figure 3). Access to the admin control of CQPweb Lancaster was granted. This allows the corpus to be indexed via CQPweb's admin control, linking each essay to its corresponding background information.

**Figure 3**
*ICNALE Written Essays in CQPweb (Lancaster) with Metadata Index (Restricted Query Page)*

As for the CQPweb version of ICNALE essays, the corpus was published with open-access rights and can be accessed via the CQPweb user interface, free of charge, by all CQP web Lancaster users at this address (https://cqpweb.lancs.ac.uk/icnale13/ accessed 15 May 2024). Compared to the official ICNALE online interface, CQPweb offers more search powers and functionalities, such as advanced keyword and collocation analyses, richer metadata and restricted searches, and faster processing time.

**Corpus Data Analysis**

Unlike other studies on LBs, our approach was top-down. O'Flynn's (2022, pp.104–155) list of frequent LBs was used in this research. On the one hand, this might be considered a handicap because it is not fully data-driven. But on the other hand, this can be considered an opportunity to examine to what extent O'Flynn's LBs can be applied. This entails a methodological difference from other studies in which LBs were collected using the n-gram function (word sequence/bundle). Conversely, O'Flynn's LBs were used to perform orthographic queries (raw text searches). We focused here on 4-word LBs. We agree with Hyland (2008), Cortes (2004), and O'Flynn (2022) that 4-word LBs perform a wider range of functions and provide a more manageable data set. LBs lower or higher than this range are not discussed here. As for the cut-off frequency, we moderately adopted O'Flynn's 10 per million words (pmw) frequency. But when an LB surfaces in comparison with another LB higher than the cut-off frequency, the LB whose frequency is lower than the cut-off point is marked by an asterisk.

When interpreting the findings, CQPweb's relative frequency – also termed 'pmw' (per million words) count – is used (Figure 4). This means the frequency count may be relative to the corpus size. For instance, the frequency of the LB 'on the other hand' is 204.618 in the Asia group. But in the EFL group, it is only 190.689.

**Figure 4**

*Relative Frequency in CQPweb Per-million Words*
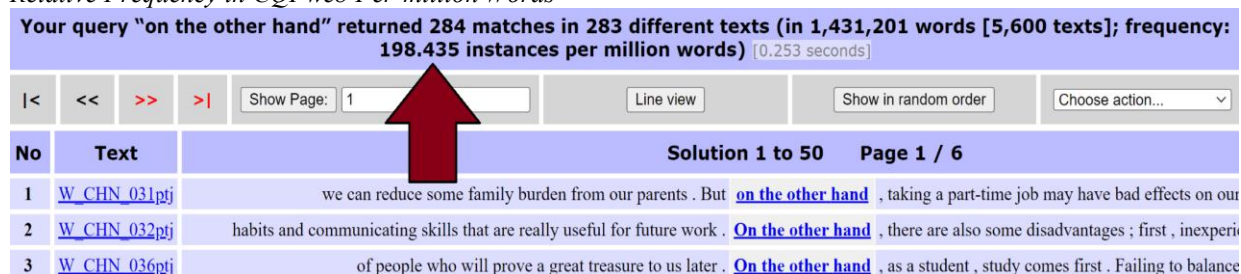


This is because the size of the sub-corpus has changed. Absolute frequency (termed a 'hit' in CQPweb) is not used in this project because the sizes of the sub-corpora are not balanced. For instance, the Japan sub-corpus (essays from students in Japan) is almost twice as large (198,949 tokens) as the Indonesia sub-corpus (103,275 tokens). The China sub-corpus is the largest (211,523 tokens). The smallest is the Hong Kong sub-corpus (51,900). The term frequency is used to refer to relative frequency throughout the paper. The goal of this study is to compare data from these countries as groups (the Asia group, consisting of ESL and EFL groups, and the ENS group). Their sizes are summarised in Figure 5.

For this reason, restricted searches were applied to particular countries relevant to the grouping. Here, we were inspired by Kachru (1990), but instead of inner, outer, and expanding circles, we used ENS, ESL, and EFL. This is because Kachru (1990) approached this from a more sociolinguistic perspective. By using more common terms (ENS, ESL, and EFL), we wish to reach not only language pedagogists but also a broader audience.

For the Asia group, data from all Asian countries were selected, as illustrated in Figure 6. For the ENS group, data from Canada, the US, the UK, Australia, and New Zealand were selected. For the EFL group, data from Indonesia, Vietnam, China, Japan, Korea, and Thailand were selected. For the ESL group, data from the Philippines, Malaysia, Hong Kong, and Pakistan were selected. Note that all countries mentioned in the ICNALE data (as of July 2024) were included.

When taking a comparative analysis of proficiency into account, searches were performed based on proficiency metadata, including restricted searches (Figure 7). In ICNALE written essays, three proficiency scales from CEFR, namely A2, B1, and B2, were present. Adapted to ICNALE metadata, the B1 category is divided into two groups: B1.1 and B1.2, with B1.2 being more proficient. Option XX refers to ENS groups. In this study, ENS is considered to have the highest proficiency level. Therefore, the (ascending) order is A2, B1.1, B1.2, B2, and ENS.

**Figure 5**
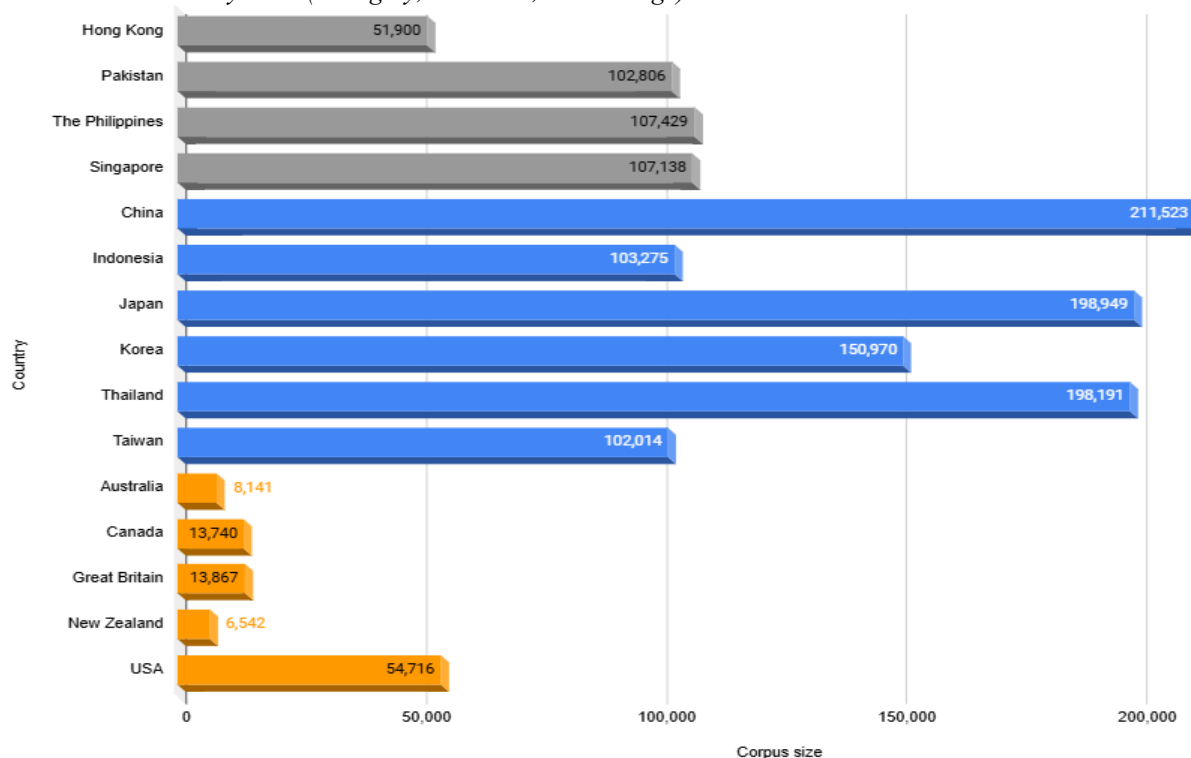*ICNALE Written Essays Size (ESL-grey, EFL-blue, ENS-orange)*



**Figure 6**
*Restricted Search by Country*



**Figure 7**
*Restricted Search by Proficiency*



For frequency counts of individual LBs in Figure 8 based on proficiency level, they were used as is. But for the frequencies of structural and functional types, LB disambiguation was applied when necessary. This is because some LBs may be structurally or functionally ambiguous (Güngör & Uysal, 2020; Salazar, 2014). This means that a given LB may be categorised as one type in a specific context, but as another type in a different context. For example, the LB 'the result(s) of' may serve as a

'stance' function in (1) or a 'referential' function in (2). This is because in (1) the LB conveys the author's stance or evaluation, whereas in (2) it refers to a specific outcome. We distributed the functional and structural categorisation task to three coders trained under the same protocol. In most cases (96%), all coders completely agreed (3 out of 3 provided the same label) on the structural and functional categorisation labels. When there was a disagreement (2 out of 3 or all three different), we inspected the results and made a decision.

(1) The results of the experiment strongly support our hypothesis (stance)
(2) The results of the study were analysed using CQPweb (referential)

**Figure 8**

*Frequency of Individual LBs in Asia (Yellow) and ENS (Red) Groups (with LL Score Indicated Before Each LB – Sorted From Highest to Lowest)*
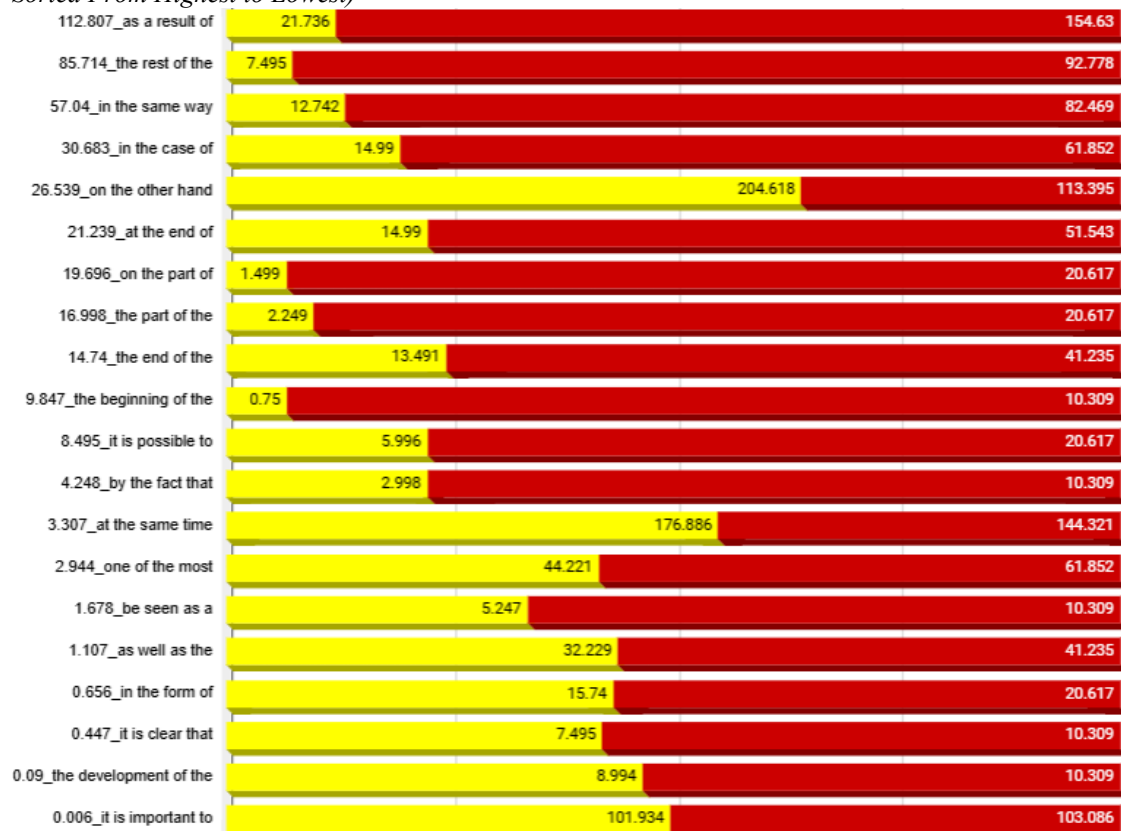


To measure differences across individual LBs or types (structural or functional), Log-Likelihood, a common measure applied in a number of previous LB studies comparing two groups – as in Hyland and Jiang (2018), Fajri et al. (2020), and Narkprom and Phoochareonsil (2022) – was used. However, when differences were calculated from more than two different groups (e.g., proficiency levels), the z-score was used. When a correlation analysis was required, the Pearson correlation coefficient was preferred, assuming the data were normally distributed (as determined by a Shapiro-Wilk test). If the assumption of normality is violated, Spearman's correlation measure is used.

Note that these findings contradict some previous LB studies. For instance, Narkprom and Phoocharoensil (2022) examined LBs in theses written by Thai and ENS groups. They found that in some sections, Thai students overused LBs more than ENS groups. They believed this was driven by expectations from Thai universities for greater criticality in those sections. A similar trend was found by Hadizadeh and Jahangirian (2022). They found that a non-ENS group (native Iranian students) used more LBs than an ENS group. They argued that this could be attributed to the training these students received, which included the use of LBs as one of its academic writing focuses.

These studies suggest that some interventions could have been applied to ensure that the LBs used by students from non-ENS groups had the same or a similar profile as those of the ENS group. From the list of the top-10 most frequent LBs in the Asia and ENS groups (see Figure 9), most were shared. LBs unique to the Asia group were 'on the one hand' and 'in the form of'. In the ENS group, LBs unique to this group were 'the rest of the' and 'in the same way'.

**Figure 9**

*Top-10 Most Frequent LBs in The Asia and ENS Groups (Shared LBs: Blue)*
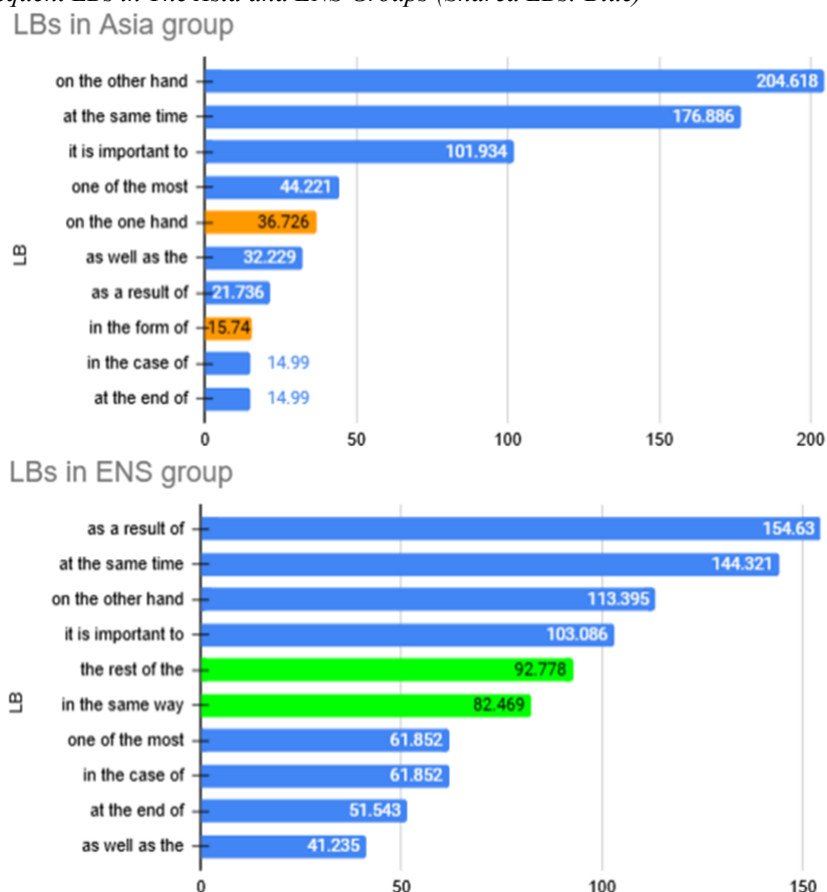


It is interesting to note that the whole correlative conjunction pair ('on the one hand' & 'on the other hand') was present in the Asia group, but in the ENS group, only 'on the other hand' was present (Figure 10 serves as an example). This suggests that a prescriptive approach claiming that 'on the one hand' and 'on the other hand' must appear hand in hand to explain contradictory ideas should be reconsidered, as this finding opposes it. This discrepancy between the frequency of use of these correlative conjunctions was also observed by Chen and Baker (2010), Fajri et al. (2020), Candarli and Jones (2019), and Hyland and Jiang (2018). Hyland and Jiang (2018) even referred to 'on the other hand' as a 'perennial favourite', with no mention of 'on the one hand'.

**Figure 10**

*Use of 'On the One Hand' and 'On the Other Hand' [W_CHN_064ptj:Asia text file] and 'On the Other Hand' Only [W_ENS_029ptj:ENS Text File]*

The Asia group was then broken down into ESL and EFL groups (see Figure 11). It can be observed that, in general, the ESL group used LBs more frequently than the EFL group. This discrepancy between the two groups was, however, lower when the two groups were aggregated with the ENS group (302 for ENS vs Asia, as shown in Table 3, and 230 for ESL vs EFL, as shown in Table 5). Similarly, but not identically, a number of studies have compared an ENS group against either an ESL group (Ibrahim & Khalit, 2020; Nekrasova-Beker & Becker, 2019; Staples et al., 2013) or an EFL group (Oktavianti & Prayogi, 2022; Uçar & Zarfsaz, 2023; Yoo & Shin, 2022). However, their data collection protocols and analytic procedures differed from those applied in this study. Appel and Murray (2023) conducted a cross-country comparison between China, Japan, and Korea. However, these countries are all EFL countries. In contrast, we argue that our study is more heterogeneous than the aforementioned studies as the data come from six ESL countries, four EFL countries, and four ENS countries.

**Figure 11**

*Frequency of Individual LBs in EFL (Yellow) and EFL (Red) Groups (With LL Score Indicated Before Each LL – Sorted from Highest to Lowest)*
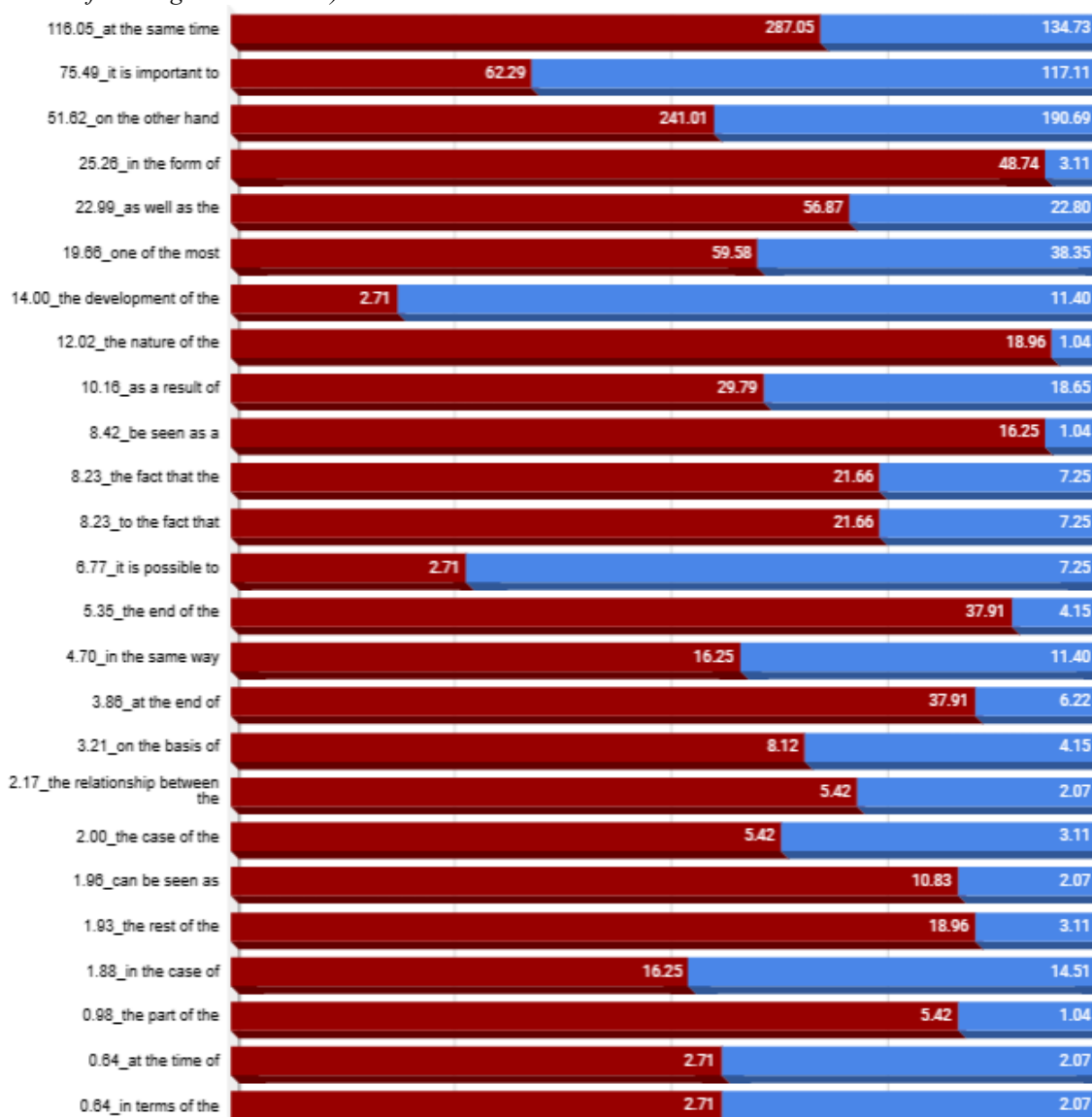


Figure 12 shows the top 10 most frequent LBs in the ESL and EFL groups. We observed fewer shared LBs as compared to the Asia vs ENS group (Table 6). For the correlative conjunction, we found that in the ESL group, 'on the one hand' was not in the top 10. Conversely, both 'on the one hand' and 'on the other hand' were in the top-10 LBs for the EFL group. This discrepancy shows that they do not always surface hand in hand, suggesting that EFL students tend to be more prescriptive than ESL and EFL groups.

**Figure 12**
*Top-10 Most Frequent LBs in the ESL and EFL Groups*
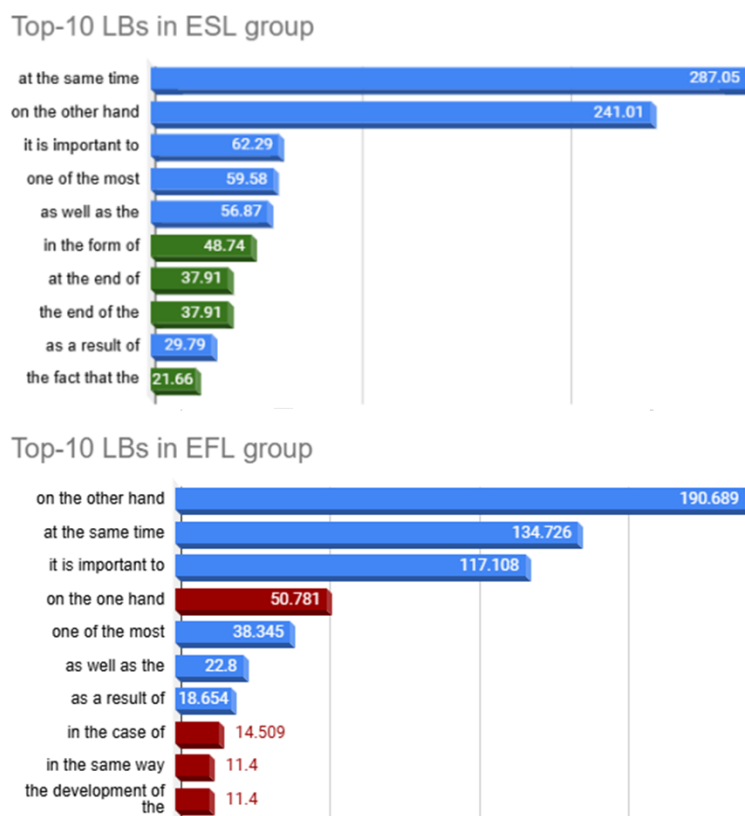


The earlier findings seem to be an indicator of proficiency: from EFL (lowest), ESL (higher), to ENS (highest). To further corroborate this, Figure 13 shows LBs organised according to the CEFR scales and ENS (as shown in the method section). There are five proficiency levels, starting from A2, increasing through B1.1, B1.2, and B2, to ENS as the highest proficiency level. If this scaling is accepted, an increase in the use of individual LBs across proficiency levels can be confirmed. This suggests a strong positive correlation (normal distribution; Pearson correlation coefficient = 0.998, *p* value = 0.0001) between the use of LBs and proficiency level.

**Figure 13**
*Overall LB Frequency Across Proficiency Levels*
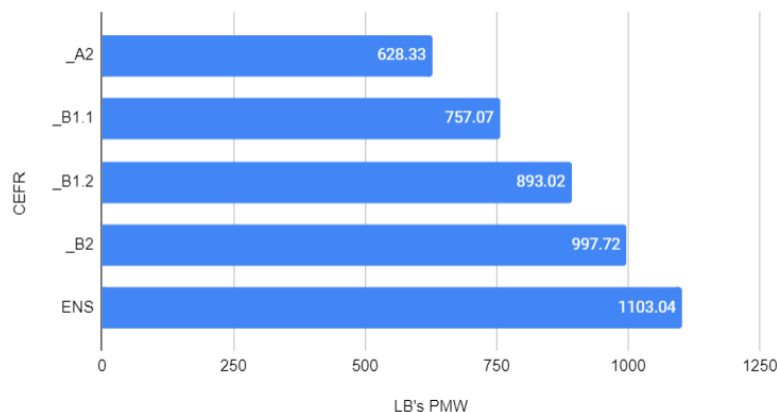


At this point, it seems that the use of LBs increases with proficiency, as previously argued by Chung and Lee (2020), who studied lexical bundles in Korean English. The same conclusion was also drawn by Li and Volvov (2018), who studied the use of LBs in email communications. A similar trend was observed by Yan (2022), who studied a corpus of written and oral exams sat by Chinese students. While other findings align with the aforementioned studies in terms of frequency, it might be worth proceeding to a qualitative evaluation, as argued by Huang (2015). This is an aspect missing from our

studies that we wish to incorporate into future research.

**Structural types**

This sub-section describes the structural categories of LBs in the Asia and ENS groups, and in the ESL and EFL groups. For the Asia vs ENS (Figure 14) groups, all structural categories were present. In terms of frequencies, more LBs (see under Greater) were used by the ENS group in all categories. The range of differences (see LL scores) was highest for NP, followed by PP and VP. As for the frequency order of structural types, for Asian learners, the (descending) order was PP, VP, NP, others; while for the ENS group, the order was PP, NP, PP, others. That PP dominates over other structural categories is to be expected because of all the LBs proposed by O'Flynn for use in this study, PP-based bundles (such as 'at the same time' as shown in Figure 15) form the majority.

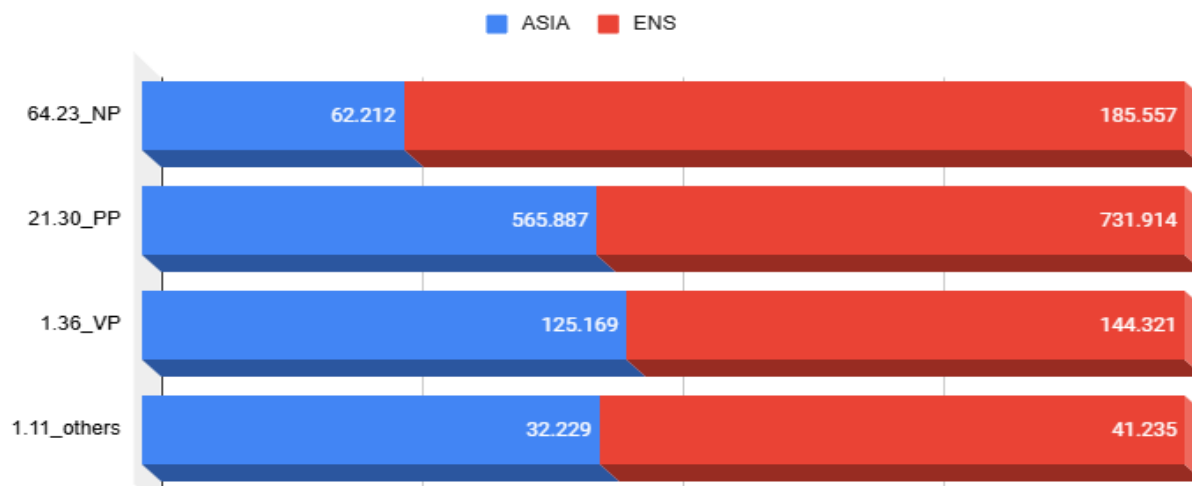**Figure 14**

*Structural Categories of LBs: Asia vs ENS*



ASIA | ENS

| | ASIA | ENS |
| --- | --- | --- |
| 64.23_NP | 62.212 | 185.557 |
| 21.30_PP | 565.887 | 731.914 |
| 1.36_VP | 125.169 | 144.321 |
| 1.11_others | 32.229 | 41.235 |

**Figure 15**

*'At the Same Time' as One of the Most Frequent PP-based LB (174.618 pmw Generated via CQPweb Concordance)*



| problem , Thai people will get lung cancer . | **On the other hand** | , people having dinner in the restaurant shoul |
| --- | --- | --- |
| e cancer has been present for 3 to 6 months . | **At the end of** | the cycle of the cancer , I saw how a regular s |
| ppiness and the atmosphere wo n't be good . | **On the other hand** | , we should let smokers know how to behave |
| ners , but is it a really " professional " skill ? | **On the other hand** | , the wages for $95 is still low for students ; a |
| t not ask for my parents , I feel very happy . | **At the same time** | , with the help of this money , I can do many |
| smoking will have more terrible influences . | **At the same time** | tobacco cost a lot of our money day after day |
| ers and mothers will sorry and be not happy | **at the same time** | . They will study hard in the next term to finis |
| b certainly brings about several advantages . | **On the one hand** | , by taking a part-time job , students can acqu |
| you may do a very bad thing to him or her . | **On the other hand** | , many smokers always leave cigarette ends a |
| camp for training and game and so on . But , | **on the other hand** | , there is a person to say to study in a college |

This finding is partially echoed by Latif et al. (2022), who compared English novels written by ENS and Pakistani groups (ESL). It was shared that the ESL group used more PP-based LBs. For the ENS group, a different observation was made. In Latif et al. (2022), VP-based bundles were used more frequently. However, our findings contradict those of Akbulut (2020) and Narkprom and Phoocharoensil (2022). Akbulut (2020), who compared ENS and non-native (Turkish) journal paper authors, found that the authors in the ENS groups used more NP-based bundles, while authors in the non-ENS group used more VP-based bundles. Narkprom and Phoocharoensil (2022), who compared dissertations written by ENS and non-ENS (Thai students) groups, observed that VP-based bundles were predominantly used. Akbulut (2020), who compared native and non-native journal paper

authors, observed that the ENS group used more NP-based bundles, while the non-ENS group used more VP-based bundles. Our findings also contradict those of Hadizadeh and Jahangirian (2022), who compared the LBs used by Iranian writers and native English writers. They discovered that both groups predominantly used NP-based LBs.

From this point, more differences in structural categories than in individual LBs (earlier section) were observed. This may be attributed to differences in corpus data. Of all the above-referenced studies, most corpora data were obtained from structured academic writing products, such as journal papers or final projects (dissertations or theses). Only Latif et al. (2022) used a corpus of non-academic data (from literary works). The corpus data for this study – while collected within an academic writing setting – were not as structured as journal articles or final

projects, in which stylesheets and rhetorics (e.g., introduction, literature review, methodology) were tightly enforced. Instead, the only controls applied to data collection in this study were topic and word limits.

As for the structural categories of LBs between the ESL and EFL (Figure 16) groups, the ESL group seemed to present more LBs (greater) in all structural categories. The difference (LL=358) was similar to that between the Asia and ENS groups (LL=387). By category, the most similar structural type was VP-based bundles, while the most different was PP-based bundles. As for the ranks of these structural types, both ESL and EFL groups showed the same frequency order: PP, NP, VP, others. This was in contrast to the Asian vs ENS groups, in which the order varied.
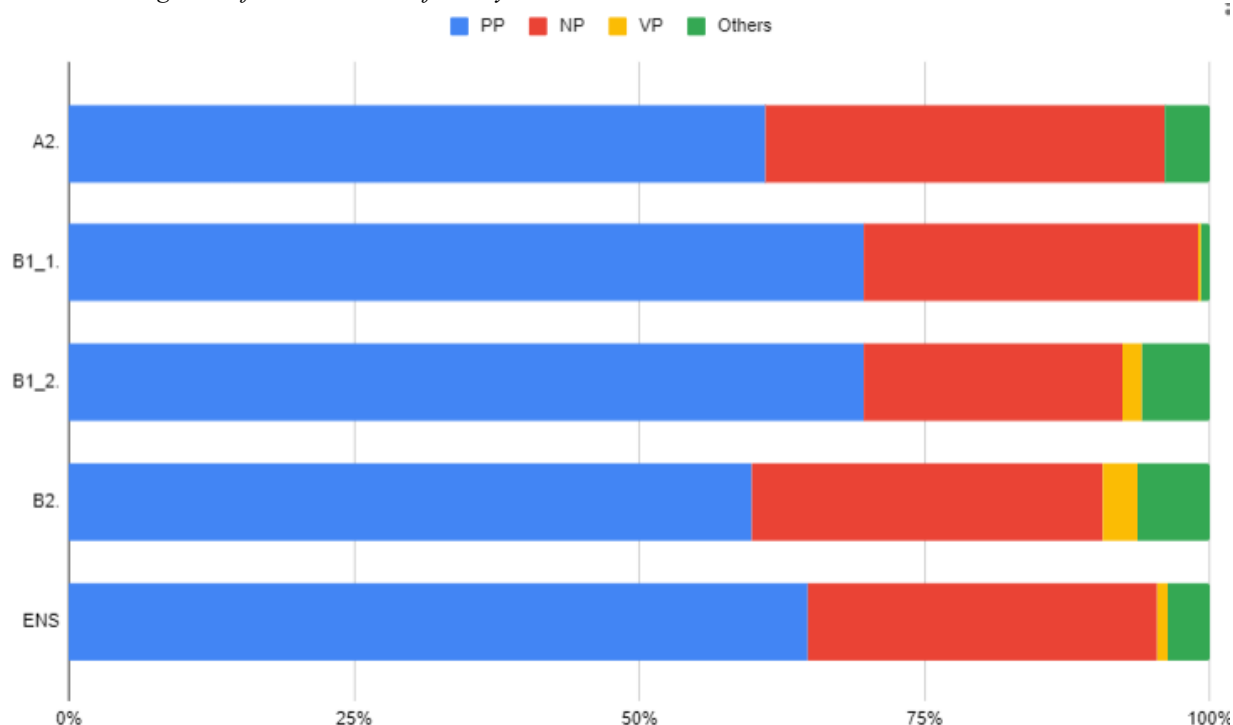
**Figure 16**
*Structural Categories of LBs: ESL vs EFL*
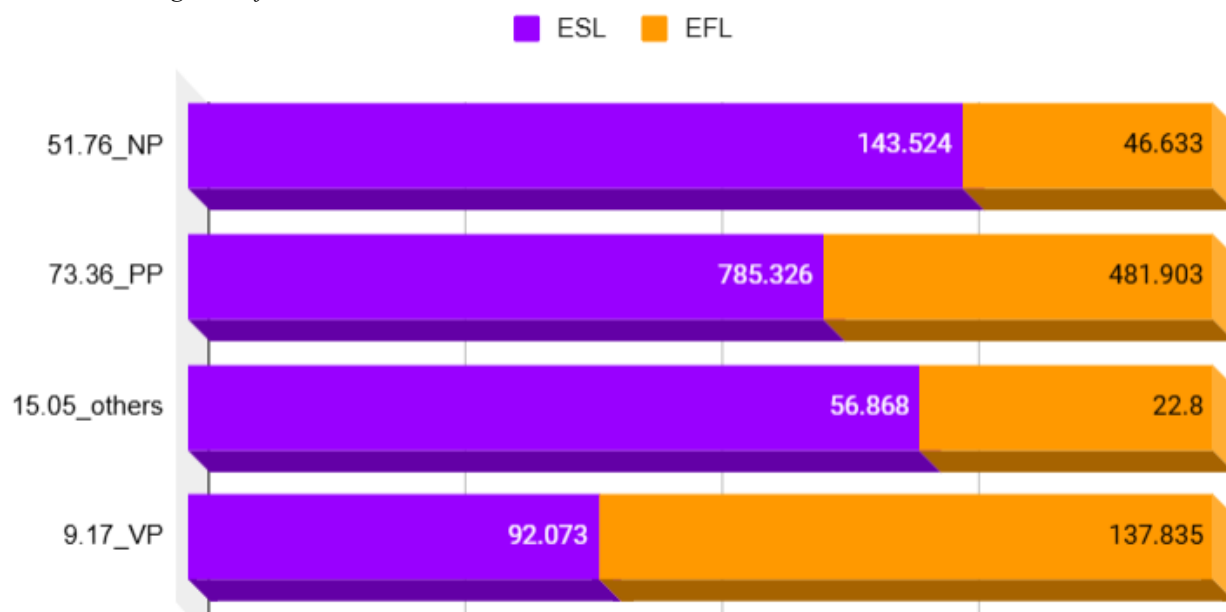


As noted earlier, no studies have specifically compared EFL and ESL. To address this omission, a comparison of two studies—Ibrahim and Khalit (2020), who studied LBs in Malaysian learner corpora, and Fajri et al.'s (2020) subset of Indonesian authors—was conducted. Indonesia and Malaysia are neighbouring countries, but in Malaysia, English is a second language, whereas in Indonesia, it is a foreign language. The dominant structural category among Malaysians was NP-based, whereas among Indonesians it was VP-based. As in the earlier section, it is argued that this difference may be attributed to the different controls

applied during corpus data collection. Ibrahim and Khalit (2020) collected argumentative essays obtained from university entrance tests, while Fajri et al. (2020) collected data from published journal papers. Conducting more LB studies (in the context of ESL vs EFL) under the same methodological control is a sensible suggestion, because it allows us to confirm to what extent they are similar or different from our findings, as well as the reasons behind the findings. Regarding structural categories across proficiency levels, the most frequent LBs were PP-based at all levels (see Figure 17).

**Figure 17**
*Structural Categories of LBs Across Proficiency Levels*



**Functional types**

All functions are present in the findings. This is because even though some LBs are absent (never used), other LBs with the same functions are present. For the Asia vs ENS groups (Figure 18), LB frequency is higher in the ENS group across all categories (except transition signals). The largest difference is in the resultative function, and the smallest difference is in the engagement function.

**Figure 18**
*LB Functional Categories of Asia and ENS Groups*
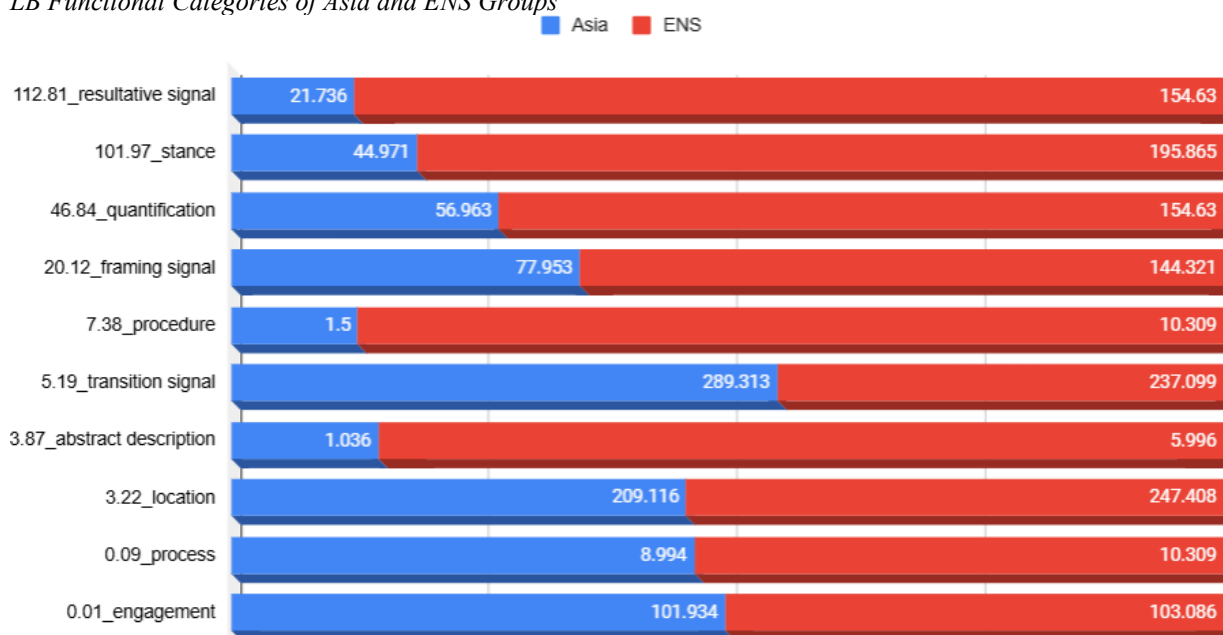


This finding partially aligns with Akbulut (2020), who compared Turkish and ENS academic authors. It echoes the finding that the largest difference was found in the resultative function, such as 'as a result of', as shown in Figure 19. However, the findings in the present study contradict those of

Fajri et al. (2020) and Hadizadeh and Jahangirian (2022). Fajri et al. (2020) compared L1 and L2 English authors and found that the greatest difference was in the procedure function, while the smallest one was in the location function. The two functions are research-oriented LBs. As Hadizadeh

and Jahangirian (2022) observed in their master's thesis, groups of Iranian and English students showed the largest difference in the structuring function, while the least difference was observed in both location and quantification LBs.

**Figure 19**

The use of 'as a result of' in one of the ENS texts



**W_ENS_119smk**

> Later studies showed , however , that there was no effect whatsoever on business performance figures , and that public opinion strongly supported the move .
>
> Today , restaurants are cleaner and more hygienic , and provide an attractive atmosphere more conducive to leisurely conversation **as a result of** the ban on smoking .
>
> Public smoking also sets a poor example for children and young people , who see smoking as a natural part of everyday activities rather than being educated in the harm caused by cigarette smoke .

Looking at the findings, factors that affected variations seem to vary. But one thing that distinguishes this study from those aforementioned is the degree of structuring of texts collected as our corpus data. In the argumentative essays collected for this research, no rhetorical structure was enforced. This is in contrast to structured academic writing products such as theses, dissertations, and journal papers used in the aforementioned studies, in which stylesheets (e.g., introduction, methodology, and literature review) had to be respected. To what extent this affects the distribution of LBs across functional types remains to be further researched. In terms of methodology, the project closest to this research is Anwar et al. (2020), who also observed ICNALE data, specifically for Pakistani students and ENS. But even in their study, the observation was

different. They found that the description function was most frequent in both Pakistani and ENS groups. The least frequent function was the transition function for the Pakistani group and the engagement function for the ENS group.

For the ESL and EFL groups (Figure 20) in this study, the same trend displayed earlier by the ENS vs Asia groups was not observed. The greatest difference was not in resultative but in location. The smallest difference was in procedure, not engagement. For ESL, there was no use of LBs in the abstract description category. We could not find a study that compared EFL and ESL using the same methodology and data analytic protocol. A number of studies that separately examined data from ESL or EFL countries exist, but their results also varied.

**Figure 20**

*LB Functional Categories of ESL and EFL Groups*
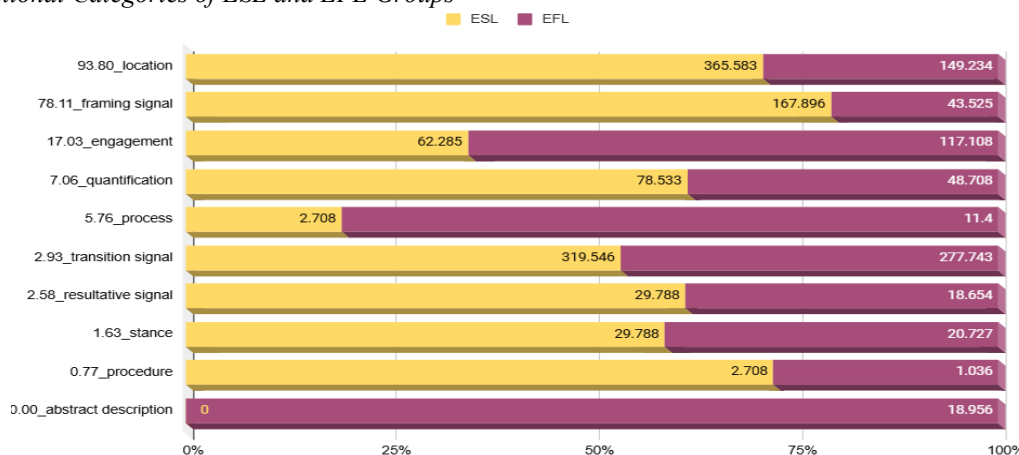


Oktavianti and Prayogi (2022), who studied native Indonesian authors, found that the most frequent function was procedure, while the least frequent were resultative and framing. This differs from our studies, in which location- and procedure-based LBs were the most frequent functional types. Beng and Yuen (2015) compared the functional

types of LBs across two domains, both by Malaysian authors. Even with two domains from the same ESL country, the results vary. While the description function was most frequent in both the science and art domains, the least frequently used function was resultative in the art domain, and engagement in the science domain. Variations in the findings of these

studies support our argument that multiple factors may be at play in the distribution of LB functions. In addition, not all studies of LBs included the observation of LB functional types (Appel & Murray, 2023; Chung & Lee, 2020; Ibrahim & Khalit, 2020). For such studies, a comparison is not possible.

When LBs were analysed by proficiency level (Figure 21), it was evident that transition and location were among the most frequent functional types across all proficiency levels. Note that for non-native proficiency levels, transition always ranked highest, followed by location. An exception was B2, where the frequencies of location and transition signal LBs were almost identical. For the ENS group, conversely, location ranked highest, while the transition signal ranked second. Process and procedure functions always ranked lowest in all groups. As for other functional types, configurations varied.

**Figure 21**

*Functional Types of LBs Across Proficiency Levels*
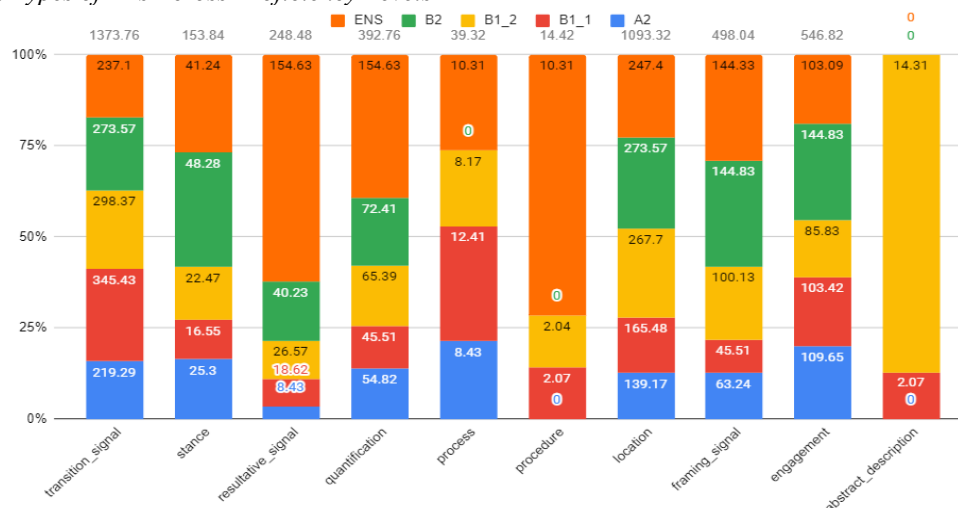


The largest differences were observed for resultative signal, quantification, and procedure (see Figure 22). This finding can characterise the LBs of native and non-native authors. The ENS group always scored higher than all proficiency levels. For instance, it scored 154.63 for resultative and quantification, while other groups only scored below 50 and 80, respectively. This finding is partially echoed by Fajri et al. (2020), who observed extremely large differences in resultative and quantification, but not in procedure. Likewise, Narkprom and Phoochaorensil (2022) observed a considerable difference in quantification, but not in procedure or resultative. Our findings contradict those of Anwar et al. (2020), in which none of these three functions showed a considerable difference.

**Figure 22**

*Differences (z-score) of Functional Type Categories Across Proficiency Levels*
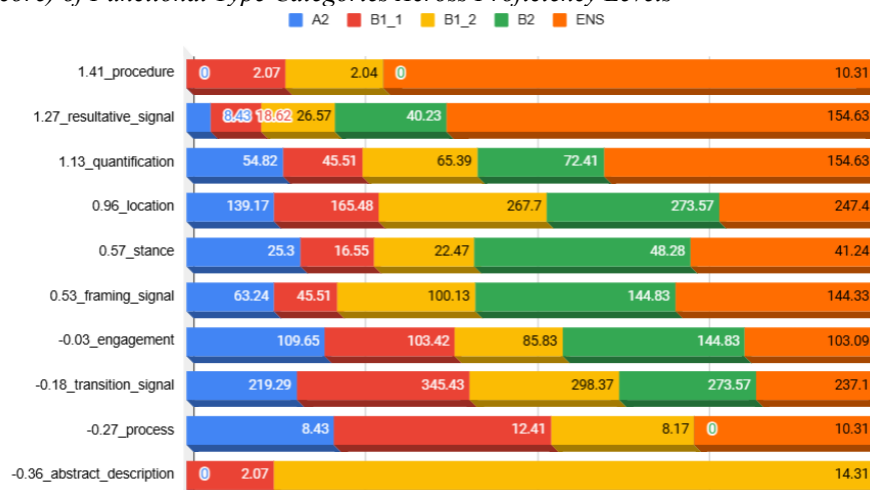
In terms of broad categories of functional types, this study shows that the most frequent LBs are text-oriented or discourse-organiser types, such as 'in terms of', as shown in Figure 23. This is true not only when comparing the Asia and ENS groups (Figure 24), but also when comparing the ESL and EFL groups (Figure 25). The order of frequency is also identical for both groups: Text-Research-Participant (descending). In a comparison of Asia and ENS groups, it is evident that functional types of LBs are more frequent across all categories. This differs slightly from the comparison between the ESL and ENS groups, in which the EFL group was greater, but only in terms of the participant category. However, on aggregate, the ENS and ESL groups still dominated over the Asia and EFL groups.

**Figure 23**
*An Example of the Use of 'In Terms of the' In an ESL Text File*



W_SIN_141smk

It is essential to implement bans on smoking in restaurants within the country , but it is not feasible and considerate to ban smoking at all restaurants .

Second-hand smoke is detrimental to non-smokers **in terms of the** human health .

Also , many non-smokers exhibit strong dislike towards the smell of the tobacco smoke .

**Figure 24**
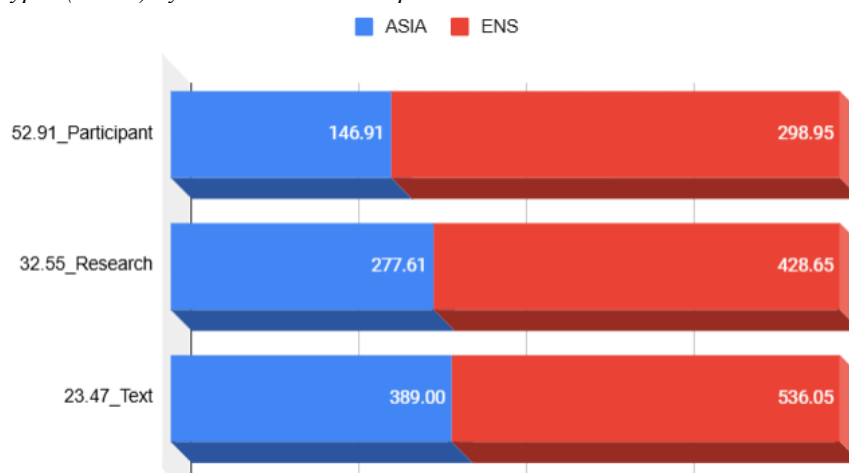*LB Functional Types (Broad) by Asia and ENS Groups*



**Figure 25**.
*LB Functional Types (broad) by ESL and EFL Groups*
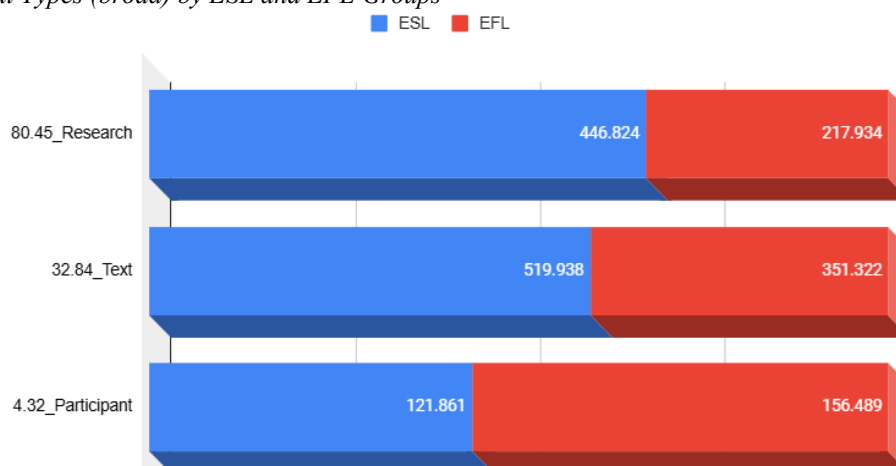


This functional type configuration is not echoed by other LB studies. In many of those, LBs were usually oriented towards the research category, as shown by Hyland (2008), who studied LBs in postgraduate published writing. The same trend was also observed by Kurniawan and Haerunissa (2023), who studied LBs in the context of published and rejected article manuscripts, and also by Syihabudding and Harjanto (2022), who studied LBs in master's theses published at an Indonesian

university. As established earlier, it is argued that a possible reason for this is differences in corpus data-collection protocols.

The data in the aforementioned studies were collected from highly structured texts, such as academic journal papers or doctoral theses. However, the corpus used in this research was collected from students' unstructured argumentative essays. But whether or not genre corresponds to the orientation of textual bundles may need further investigation, because Guan's (2022) study, whose object was also argumentative essays like this present study, notes that referential LBs are the most frequently used function, followed by stance and discourse organisers. If we follow Candarli and Jones (2019), this means their essays are more research-oriented than text-oriented, even when the same control is applied to the data. This may be attributed to the heterogeneity of Asian learners in the corpus data for this project. However, more comprehensive studies are needed to understand the extent to which the findings for broad categories can be generalised.

In terms of orientation and proficiency level (Figure 26), we observed a similar trend to the broad categories expressed earlier in text-orientation LBs. This, to some extent, stands in contrast to Guan (2022), who discovered that lower proficiency L2 authors use stance and discourse organisers LBs in argumentative essays, which correspond to Hyland's participant and text-orientation LBs. Note that he argues for the frequent use of research-oriented LBs. This has not been confirmed in this study. All proficiency groups exhibited a similar trend. While not the most frequent LBs, the proportion of research-oriented LBs fluctuated across proficiency levels, from 30–35%.

In another study, Matte and Goulart (2020) explain that beginner-level students used more LBs related to concrete references, while intermediate-level students used more textual organisation and stance. In this research, this was only true for intermediate-level students (assuming B1 and B2 are intermediate). For A2 and ENS, we did not observe the same trend. We therefore suggest that ICNALE also covers beginner-level student data (A1) and proficient-level student data (C1 and C2), so that the extent to which they align with the aforementioned findings can be confirmed.

**Figure 26**

*Orientation of LBs Across Proficiency Levels*
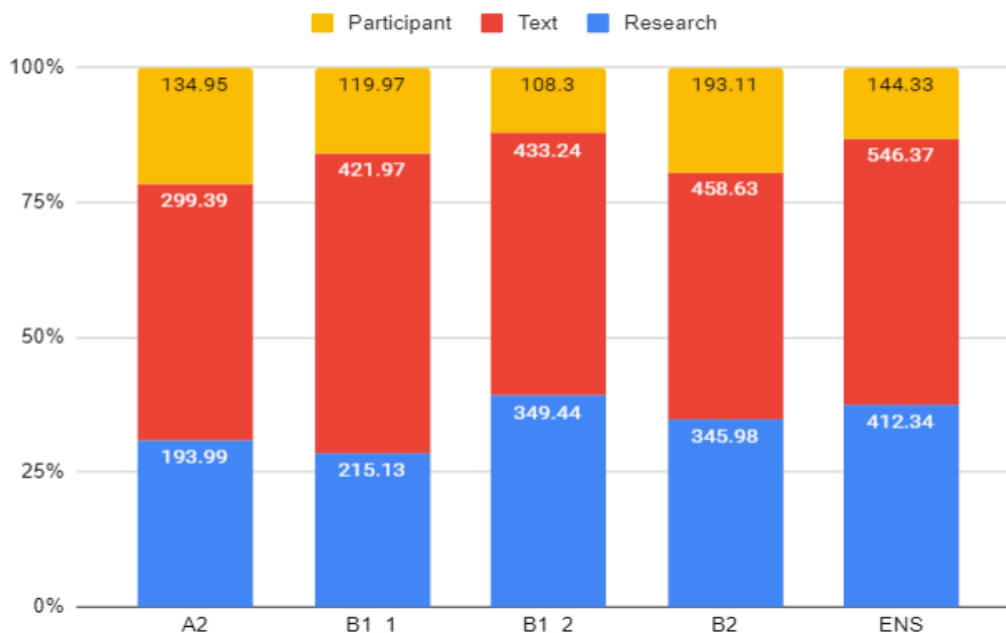


For broad categories of functional types, the trajectory is very clear. The more proficient the authors are, the more frequently they use LBs. This applies consistently to all three categories. As for the magnitude of differences, all three types exhibited different levels (Figure 27). The (descending) order was participant-text-research. Conversely, the order in Fajri et al. (2020) is text-research-participant, while in Narkprom and Phoocaorensil (2022), it is research-participant-text. All three studies exhibit different patterns in the degree of difference.

**Figure 27**
*Differences in LB's Orientation Across Proficiency Levels*
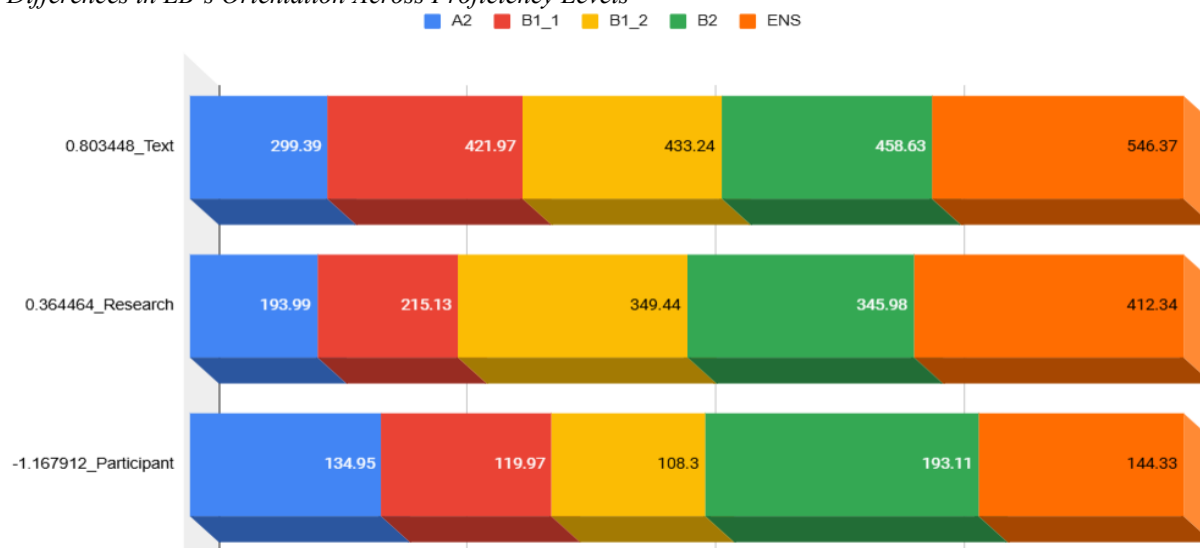


## DISCUSSION

Our findings, expressed earlier, have pedagogical and theoretical implications. The observed imbalance, in which text-oriented bundles dominate stance and engagement bundles, indicates a need to improve persuasion and argumentation. In terms of materials development, stance (e.g., *it is important to note that* this suggests that) and engagement (e.g., we can see that, let us consider) LBs can be explicitly incorporated into textbooks or digital learning platforms. This helps students demonstrate their mastery of authorial and rhetorical skills (Hyland, 2008; O'Flynn, 2022).

Second, proficiency levels may be used to determine which types of LBs to introduce. For instance, more text-oriented LBs could be introduced to beginner students, while intermediate or advanced students need to be exposed to more stance and engagement LBs. The aim is for EFL and ESL students to progress and match the LB profile of ENS, including the diversity of LBs (Ädel and Erman, 2012; Salazar, 2014).

Third, classroom activities should see both identification and production as equally important. For example, teachers can start with corpus-informed exercises by asking students to highlight LBs in model texts, compare LBs across proficiency levels, and replace odd LBs or LBs that are influenced by students' native language with targeted LBs (Cortes, 2004). Where personalised feedback is hard to provide (perhaps due to large classes), students can resort to automated tools containing an LBs feature, such as the one developed by O'Flynn (2022).

While essays produced in ICNALE are, as the control expresses, quite generic, LB profiles in disciplinary writing, as shown in Kwary et al. (2017), may differ. This aligns with Duff (2020), who highlighted that writing development is inseparable from enculturation into disciplinary discourse. Overall, this situates lexical bundle use as a marker of both linguistic proficiency and socialisation into academic norms. This provides a useful bridge between corpus-driven accounts of learner language and broader theories of discourse competence.

## CONCLUSION

This study has fulfilled its primary aim, which is to provide an overview of LBs among Asian learners of English in EFL and ESL contexts, compared with an ENS group. A significant positive relationship between the frequency of individual LBs and proficiency level is evident, echoing previous studies that LBs may correlate with proficiency levels. In terms of structural types, in all three groups (and for the majority of LBs), the most common structural type is text-oriented LBs, which stands in contrast to previous LB studies in which research-oriented LBs were the most frequent. It is very likely that the key to this variation is the data type. Previous studies mostly covered highly structured, research-based texts in which standardised stylesheets and rhetoric were strongly enforced, such as theses, dissertations, and peer-reviewed, published academic journal papers. Conversely, the corpus for this research was derived from argumentative and unstructured essays, not from research-based or published work. In conclusion, this research demonstrates that LB usage among English learners is context-sensitive and offers valuable insights into language development and writing structures reflecting English-language development in Asia, across ESL and EFL groups compared to an ENS group.

Note that we acknowledge some limitations in our corpus and methods. First, the controls applied to

ICNALE data collection can limit the analytic scope; we intend to address this in the future by expanding ICNALE controls, as in MICASE (Simpson-Vlach & Leicher, 2006) or MICUSP (Römer & O'Donnell, 2011). Second, we relied on O'Flynn's (2022) top-down LBs, which, while systematic on the one hand, can potentially leave some important LBs used by learners undetected on the other. Future studies on LBs in this context can still be improved by expanding research data coverage or the analytical variables. It is recommended to incorporate corpus data from published works, as this research solely focuses on unstructured argumentative essays. In addition, more varieties of academic works can be added, such as response papers, term papers, proposals, or reports. Incorporating proficiency levels absent from this study -A1, C1, and C2- is a worthwhile avenue to pursue. It is also recommended to conduct cross-domain analytical LB studies (e.g., history, legal, engineering, life sciences, economy, language, and art), an analytical variable absent from this study. It is also recommended to combine both top-down and bottom-up approaches. By doing so, comparative profiles between structured and unstructured texts may be obtained.

## ACKNOWLEDGEMENT

## COMPETING INTEREST

The authors declare that they have no competing interests, whether financial or non-financial, that could have influenced the work reported in this manuscript.

## REFERENCES

Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundle approach. *English for Specific Purposes*, *31*(2), 81–92. https://doi.org/10.1016/j.esp.2011.08.004

Akbulut, F.D. (2020). A bibliometric analysis of lexical bundles usage in native and non-native academic writing. *Dil ve Dilbilimi Çalışmaları Dergisi*, *16*(3), 1146–1166. https://doi.org/10.17263/jlls.803583

Anwar, N., Malik, M., & Samhadi, J. (2020). Lexical bundles in academic discourse: A comparative study of functional use of lexical bundles in native and non-native learner corpora. *Epistemology*, *7*(3), 28–34. https://journal.epistemology.pk/index.php/epistemology/article/view/152

Appel, R., & Murray, L. (2023). A contrastive interlanguage analysis of lexical bundles in English as a foreign language writing: L1 Chinese, Japanese, and Korean. *JALT Journal*, *45*(1), 35–58. https://doi.org/10.37546/jaltjj45.1-2

Beng, C. O. S., & Yuen, C. K. (2015). Functional types of lexical bundles in reading texts of Malaysian university English test: A corpus study. *GEMA Online Journal of Language Studies, 15*(01), 77–90. https://doi.org/10.17576/GEMA-2015-1501-05

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*(3), 371–405. https://doi.org/10.1093/applin/25.3.371

Biber, D., Johansson, S., Leech, G. N., Conrad, S., & Finegan, E. (2021). *Grammar of spoken and written English*. John Benjamins Publishing Company. https://doi.org/10.1075/z.232

Budiwiyanto, A. (2023). Use of lexical bundles in an online comprehensive dictionary of Indonesian (KBBI Daring). *Lexicography*, *10*(1), 4–22. https://doi.org/10.1558/lexi.25177

Candarli, D., & Jones, S. (2019). Paradigmatic influences on lexical bundles in research articles in the discipline of education. *Corpora*, *14*(2), 237–263. https://doi.org/10.3366/cor.2019.0170

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, *14(2)*, 30–49. http://dx.doi.org/10125/44213

Chenda, S., Safnil, S., & Syafryadin, S. (2022). Stance expressions in the introduction of English research articles written by Cambodian authors. *Register Journal*, *15*(2), 183–200. https://doi.org/10.18326/rgt.v15i2.183-200

Chung, J., & Lee, K. R. (2020). Structural analysis of lexical-bundles in uncontrolled environment of spoken discourse: The same heritage language groups. *The Journal of AsiaTEFL*, *17*(2), 479–492. https://doi.org/10.18823/asiatefl.2020.17.2.11.479

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397–423. https://doi.org/10.1016/j.esp.2003.12.001

Estaji, M., & Montazeri, M. R. (2022). Native English and non-native authors' utilisation of lexical bundles: A corpus-based study of scholarly public health papers. *Southern African Linguistics and Applied Language Studies*, 40(2), 177–199. https://doi.org/10.2989/16073614.2022.2043169

Duff, P. A. (2010). Language socialisation into academic discourse communities. *Annual Review of Applied Linguistics, 30*, 169–192. https://doi.org/10.1017/S0267190510000048

Fajri, M. S. A., Kirana, A. W., & Putri, C. I. K. (2020). Lexical bundles of L1 and L2 english professional scholars: A contrastive corpus-driven study on applied linguistics research articles. *Journal of Language and Education*, 6(4), 76–89. https://doi.org/10.17323/jle.2020.10719

Faqih, M. S., & Harjanto, I. (2022). English lexical bundles in the graduate theses: The frequency, structure and distribution. *JEELS (Journal of English Education and Linguistics Studies)*, 9(1), 27–49. https://doi.org/10.30762/jeels.v9i1.3652

Fitriati, S. W., & Wahyuni, S. (2019). Lexical bundles in WhatsApp conversation between native and non-native English speakers. *Proceedings of the UNNES International Conference on English Language Teaching, Literature, and Translation (ELTLT 2018)*, 188. Atlantis Press. https://doi.org/10.2991/eltlt-18.2019.61

Flowerdew, J., & Mahlberg, M. (Eds.). (2009). *Lexical cohesion and corpus linguistics* (Vol. 17). John Benjamins Publishing Company. https://doi.org/10.1075/bct.17

Guan, J. (2022). A contrastive study on lexical bundles in argumentative writing by L1-Chinese and L1-English undergraduates. *English Language Teaching*, 15(12), 43. https://doi.org/10.5539/elt.v15n12p43

Güngör, F., & Uysal, H. H. (2020). Lexical bundle use and crosslinguistic influence in academic texts. *Lingua*, 242, 102859. https://doi.org/10.1016/j.lingua.2020.102859

Hadizadeh, A., & Jahangirian, S. (2022). Lexical bundles and disciplinary variation in master theses. *Language Teaching and Educational Research*, 5(2), 59–79. https://doi.org/10.35207/later.1152493

Hardie, A. (2012). CQPweb—Combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17(3), 380–409. https://doi.org/10.1075/ijcl.17.3.04har

Hardie, A. (2024). *CQPWeb Lancaster* (v.3.3.18) [Computer software]. https://cqpweb.lancs.ac.uk/ (Accessed 2024-05-27)

Huang, K. (2015). More does not mean better: frequency and accuracy analysis of lexical bundles in Chinese EFL learners' essay writing. *System*, 53, 13–23. https://doi.org/10.1016/j.system.2015.06.011

Hyland, K. (2008). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41–62. https://doi.org/10.1111/j.1473-4192.2008.00178.x

Hyland, K., & Bondi, M. (Eds.). (2006). *Academic discourse across disciplines*. Peter Lang. https://doi.org/10.3726/978-3-0351-0446-2

Hyland, K., & Jiang, F. (2018). Academic lexical bundles: How are they changing? *International Journal of Corpus Linguistics*, 23(4), 383–407. https://doi.org/10.1075/ijcl.17080.hyl

Ibrahim, E. H. E., & Khalit, F. F. (2020). Lexical bundles in Malaysian ESL learner corpora. *International Journal of Academic Research in Business and Social Sciences*, 10(14), Pages 23-30. https://doi.org/10.6007/IJARBSS/v10-i14/7359

Ishikawa, S. (2023). *The ICNALE guide: An introduction to a learner corpus study on Asian learners' L2 English* (1st ed.). Routledge. https://doi.org/10.4324/9781003252528

Kachru, B. B. (1990). World Englishes and applied linguistics. *World Englishes*, 9(1), 3–20. https://doi.org/10.1111/j.1467-971X.1990.tb00683.x

Kang, S., Shin, Y. K., & Yoo, I. W. (2024). Using lexical bundles to teach prepositions to Korean EFL students: Corpus-based instructed SLA. *Journal of Second Language Studies*, 7(1), 75–98. https://doi.org/10.1075/jsls.00022.kan

Kilgarriff, A., Rychlý, P., Jakubíček, M., Kovář, V., Suchomel, V., Michelfeit, J., Herman, Svoboda, T., Blahuš, M., Mikušek, O., & Kubík, M. (2024). *Sketch Engine v.2.36.5* [Computer software]. Lexical Computing. https://www.sketchengine.eu/ (Accessed 2024-05-27)

Kilgarriff, A., Rychly, P., Smrž, P., & Tugwell, D. (2008). The Sketch Engine. In T. Fontenelle (Ed.), *Practical lexicography* (pp. 297–306). Oxford University Press. https://doi.org/10.1093/oso/9780199292332.003.0020

Kim, Y. (2009). Korean lexical bundles in conversation and academic texts. *Corpora*, 4(2), 135–165. https://doi.org/10.3366/E1749503209000288

Kurniawan, E., & Haerunisa, Z. F. (2023). A comparative study of lexical bundles in accepted and rejected applied linguistic research article introductions. *Studies in English Language and Education*, *10*(2), 628–648. https://doi.org/10.24815/siele.v10i2.28119

Kwary, D. A., Ratri, D., & Artha, A. F. (2017). Lexical bundles in journal articles across academic disciplines. *Indonesian Journal of Applied Linguistics*, *7*(1), 131. https://doi.org/10.17509/ijal.v7i1.6866

Latif, F., Dodhy, S., Tajammul, M. (2022). A comparative analysis of lexical bundles used by native and non-native novel writers. *Pakistan Journal of Social Research*, *04*(02), 1216–1231. https://doi.org/10.52567/pjsr.v4i2.618

Li, L., Franken, M., & Wu, S. (2024). Sentence initial lexical bundles in Chinese and New Zealand PhD theses in the discipline of general and applied linguistics. *Australian Review of Applied Linguistics*, *47*(1), 101–122. https://doi.org/10.1075/aral.21018.li

Li, Z., & Volkov, A. (2018). "To whom it may concern": A study on the use of lexical bundles in email writing tasks in an English proficiency test. *TESL Canada Journal*, *34*(3), 54–75. https://doi.org/10.18806/tesl.v34i3.1273

Libben, G. (2022, 20 April). Lexical representations in language processing. *Oxford Research Encyclopedia of Linguistics*. Retrieved 24 May. 2024, from https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-398.

Liu, C. Y., & Chen, H. J. H. (2020). Functional variation of lexical bundles in academic lectures and TED talks. *Register Studies*, *2*(2), 176–208. https://doi.org/10.1075/rs.18003.liu

Matte, M. L., & Goulart, L. (2020). Lexical bundles across levels of proficiency in Portuguese as a second language: an examination of bundle function. *Letras de Hoje*, *55*(4), 477-495. e38377. https://doi.org/10.15448/1984-7726.2020.4.38377

Narkprom, N., & Phoocharoensil, S. (2022). Lexical bundles in native English speakers' and Thai writers' dissertations. *GEMA Online® Journal of Language Studies, 22*(3), 43–62. https://doi.org/10.17576/gema-2022-2203-03

Nekrasova-Beker, T., & Becker, A. (2019). Lexical bundles in university course materials: From academic English to pathway to mainstream engineering. *International Journal of Corpus Linguistics*, *24*(2), 143–168. https://doi.org/10.1075/ijcl.17075.nek

O'Flynn, J. (2022). Lexical bundles in the academic writing of the arts and humanities: From corpus to CALL. *Yearbook of Phraseology*, *13*(1), 81–108. https://doi.org/10.1515/phras-2022-0006

Oktavianti, I. N., & Prayogi, I. (2022). Discourse functions of lexical bundles in Indonesian EFL learners' argumentative essays: A corpus study. *Studies in English Language and Education*, *9*(2), 761–783. https://doi.org/10.24815/siele.v9i2.23995

Prihantoro, P., & Gillings, M. (2025). The language of justice: Examining courtroom discourse in an electoral conflict. *International Journal for the Semiotics of Law – Revue internationale de Sémiotique juridique*, *38*, 2385-2407. https://doi.org/10.1007/s11196-025-10299-4

Prihantoro, P., Ishikawa, S., Liu, T., Fadli, Z. A., Rini, E. I. H. A. N., & Kepirianto, C. (2025). DICO-JALF v.1.0: Diponegoro Corpus of Japanese learners as a foreign language in Indonesia with AI error annotation and human supervision. *Jurnal Arbitrer*, *12*(3), 274–288. https://doi.org/10.25077/ar.12.3.274-288.2025

Qi, H., & Pan, F. (2020). Lexical bundle variation across moves in abstracts of medical research articles. *Southern African Linguistics and Applied Language Studies*, *38*(2), 109–128. https://doi.org/10.2989/16073614.2020.1763814

Rodrigues, A. G. (2020). Corpus Linguistics software: Understanding their usages and delivering two new tools. [Doctoral Thesis, Lancaster University]. Lancaster University. https://doi.org/10.17635/lancaster/thesis/1165 (Accessed 2024-05-24)

Römer, U., & O'Donnell, M. B. (2011). From student hard drive to web corpus (part 1): the design, compilation and genre classification of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, *6*(2), 159–177. https://doi.org/10.3366/COR.2011.0011

Salazar, D. (2014). *Lexical bundles in native and non-native scientific writing: applying a corpus-based study to language teaching* (Vol. 65). John Benjamins Publishing Company. https://doi.org/10.1075/scl.65

Shin, Y. (2018). The construction of English lexical bundles in context by native and non-native freshman university students. *English Teaching, 73*(3), 115–139. https://doi.org/10.15858/engtea.73.3.201809.115

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: new methods in phraseology research. *Applied Linguistics*, *31*(4), 487–512. https://doi.org/10.1093/applin/amp058

Simpson-Vlach, R. C., & Leicher, S. (2006). *The MICASE Handbook: A resource for users of the Michigan Corpus of Academic Spoken English*. https://www.press.umich.edu/101203

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*(3), 214–225. https://doi.org/10.1016/j.jeap.2013.05.002

Tománková, V. (2016). Lexical bundles in legal texts corpora – Selection, classification and pedagogical omplications. *Discourse and Interaction*, *9*(2), 75. https://doi.org/10.5817/di2016-2-75

Uçar, S., & Zarfsaz, E. (2023). A corpus-based teaching of lexical bundles to enhance writing skills of prospective Turkish EFL teachers. *International Journal of Learning and Teaching*, *15*(1), 42–54. https://doi.org/10.18844/ijlt.v15i1.8577

Yakut, İ., & Yuvayapan, F. (2022). Lexical bundles across disciplines: the case of research articles in the social sciences. *İnönü University International Journal of Social Sciences (INIJOSS)*, *11*(2), 272-291. https://doi.org/10.54282/inijoss.1143817

Yan, H. (2022). I think we should…: Investigating lexical bundle use in the speech of English learners across proficiency levels. *International Journal of Translation, Interpretation, and Applied Linguistics*, *1*(2), 1–16. https://doi.org/10.4018/IJTIAL.2019070105

Yoo, I. W., & Shin, Y. K. (2022). English lexical bundles in a learner corpus of argumentative essays written by Korean university students. *Corpora*, *17*(Supplement), 23–42. https://doi.org/10.3366/cor.2022.0245

Yu, J. (2023). Lexical bundle distributions across two learner corpora. *Journal of Education and Teaching*, *1*(3), 13–25. https://doi.org/10.59825/jet.2023.1.3.13

Zahra, T., Hussain, G., & Abbas, A. (2021). Discourse functions of lexical bundles in Pakistani chemistry and physics textbooks. *GEMA Online® Journal of Language Studies*, *21*(1), 221–238. https://doi.org/10.17576/gema-2021-2101-1310.37546/JALTJJ45.1-2