

An Exploratory Analysis of Linking Adverbials Used by Filipino, Pakistani, and Thai Writers of English

Randy Appel

r_appel@waseda.jp
Global Education Center, Waseda University, Japan

Corin Golding*

coringolding@aoni.waseda.jp
Global Education Center, Waseda University, Japan

Abstract

The current study provides a Contrastive Interlanguage Analysis (CIA) of linking adverbials (e.g., *furthermore*, *in conclusion*, *on the other hand*) in the second language (L2) English academic writing of post-secondary students from three countries: the Philippines, Pakistan, and Thailand. This analysis makes use of 80 essays from each of these three first language (L1) groups by way of data sourced from the International Corpus Network of Asian Learner English (ICNALE); we eschew the use of a native speaker control group in response to recent critiques of the native speaker fallacy. Quantitative and qualitative analyses revealed several noteworthy production tendencies which distinguish each English variety. These include a generally low frequency of linking adverbial tokens by Filipino writers of English, as well as a comparatively narrow range of linking adverbial types by Pakistani writers of English. In terms of functional category differences, Thai writers displayed a relatively high frequency of listing devices while Pakistani writers showed a low frequency of appositional linking adverbials, and a high frequency of resultative linking adverbials. Methodological and pedagogical implications of these findings are discussed.

Keywords: L2 English writing; linking adverbials; L1-related differences

1. Introduction

In recent years, the distinction and value of various non-native varieties of English, which include forms of English used by communities as a lingua franca or an additional language, have drawn increasing attention. Studies in this area often seek to uncover the characteristics that distinguish each target variety (e.g., Appel & Murray, 2023; Chapwanya & Nel, 2023; Shin, 2019). Early studies of this nature often took a deficit-based approach in which second language (L2) and first language (L1) writers were compared in order to highlight potential L2 production tendencies that might require teaching interventions (i.e., deficient forms/production patterns that reveal a linguistic gap lagging behind the ‘native-speaker ideal’). However, this focus has gradually expanded to include comparisons among L2 groups, with the goal of better understanding the interlanguage production tendencies that distinguish specific L2 writer groups (e.g., Appel & Szeib, 2018; Lu & Ai, 2015).

One major linguistic feature that is increasingly used to highlight differences in target English varieties is linking adverbials (LAs). LAs can be defined as single and multiword cohesive links between sections of discourse, often at or above the sentence level. As such, LAs help readers interpret text that follows in light of what has already been presented. Common examples of these devices include expressions such as *on the other hand*, *however*, and *in addition*. Given that these expressions play an important role in textual cohesion and comprehension (Biber et al., 1999), and considering they frequently pose difficulties for L2 English writers (Crewe, 1990), there has been a surge in research specifically focused on how LAs are used by L2 English users from various L1 backgrounds (e.g., Chen, 2006; Ha, 2016; Wang, 2022).

This study expands upon examinations into the use of LAs in L2 English academic text by analysing L2 English argumentative essays written by individuals from three relatively understudied countries: the Philippines, Pakistan, and Thailand. While many previous studies have used comparisons of L2 English with native speaker data as a benchmark (e.g., Appel, 2020a; Bolton et. al., 2002), this research veers away from the problematic concept of a ‘native-speaker ideal’ and the issue of which particular variety should be given prominence. Instead, we seek to better understand L2 English varieties in relation to other L2 English communities. In addition, we strive to improve upon methodological limitations frequently present in prior research of this kind by using more closely comparable corpora and larger sample sizes. To achieve these goals, this

study employs a Contrastive Interlanguage Analysis (CIA; Granger, 2015), drawing from 240 essays sourced from the International Corpus Network of Asian Learners of English (ICNALE).

2. Literature Review

2.1 Linking Adverbials (LA)

The term LA is used in the current study, although alternative naming conventions (e.g., logical connectors, sentence linking devices, discourse markers) have been used in previous studies. While these labels are often used interchangeably, or with substantial overlap in core concepts, we follow Liu (2008) in adopting the term LA. Liu (2008) argues that this term offers greater distinction from alternative cohesive devices that include inherent syntactic functions (e.g., subordinators, coordinators). In other words, an important defining feature of Las, as applied in this study, is that they play a purely semantic role in discourse cohesion and can therefore be removed from the sentences in which they are found without impacting grammatical acceptability. Examples of items falling within this category include *however, furthermore, therefore, and in addition*.

2.2 Previous Research on Las in L2 English Writing

Previous scholars have increasingly focused on the use of LAs by L2 English users, often adopting a deficit-based perspective. This viewpoint positions L2 English learners as requiring additional instruction to better imitate L1 English norms. Studies of this kind date back to at least the early 1990s (e.g., Granger & Tyson, 1996; Milton & Tsang, 1993) coinciding with the broader growth in corpus-based and corpus-driven analyses of L2 English writing (Granger, 1998). While this body of research has investigated the language produced by several non-native English writer populations, the majority of these studies have targeted one particular L2 group: L1 Chinese writers of English (e.g., Bolton et al., 2003; Hsu et al., 2020; Milton & Tsang, 1993; Yeung, 2009). The focus on these writers is, at least partially, a result of the sheer amount of data available, as L1 Chinese users of English represent one of the largest groups of non-native English users in the world. Consequently, data collection is often more manageable compared to other L1 groups with smaller populations and associated data sets available for research. Representative studies targeting L1 Chinese users of L2 English include Bolton et al. (2002) who found general LA overuse among these writers, highlighting several high-frequency items as distinctive of L1 Chinese English as a Foreign Language (EFL) writing (e.g., *so, but, and, also*). In two follow-up

studies, these overuse tendencies were largely confirmed, with the additional finding that the frequent use of ‘besides’ appears to be a distinguishing characteristic of L1 Chinese EFL writers (e.g., Chen, 2006; Lei, 2012).

While the focus on L1 Chinese writers has largely persisted, studies examining other L1 groups have begun to increase, albeit in limited numbers. For example, Arabic (Appel, 2020a; Modhish, 2012), Spanish (Carrio-Pastor, 2013; Faya-Cerquerio & Macho-Harrison, 2022), and Korean (Appel, 2020b; Ha, 2016) have all been the focus of research in recent years. Although this broadened scope represents a positive shift towards a more comprehensive understanding of L2 English production tendencies across a diverse range of L1 groups, the selection of target L1s still remains limited. As a result, a more extensive range of target language varieties should be investigated to foster a more nuanced understanding of the distinctive production tendencies in various L2 English communities, particularly those that have been underrepresented thus far.

From a theoretical standpoint, studies using a CIA approach to identify interlanguage differences can be seen as an attempt to improve our understanding of the processes underpinning the acquisition of additional languages (Granger, 2015). This perspective also ties into the potential identification of cross-linguistic influence, in which a user’s L1 impacts the production and comprehension of any subsequent languages. Thus, comparisons of L2 varieties can contribute to theoretical understandings of these important areas, whilst also providing insights that can lead to more targeted pedagogical interventions that better address the needs of specific groups of L2 learners.

2.3 Limitations in Previous L2-English, LA-based research

The current study aims to expand the range of EFL varieties examined in research of this kind while also presenting a shift away from the traditional deficit-based view of L2 English writing by forgoing the use of an L1 English reference corpus. This approach requires us to eschew the terms *over-/underuse* (except when directly quoting authors who use them) in favour of the more neutral terms *high/low frequency*, relative to other groups in this study. We assert that the latter terminology retains a focus on opportunities for pedagogical intervention without resorting to a normative data set. At the same time, we sidestep issues regarding the problematic classification of ‘native speaker’ and the challenge of accounting for differences within such a group (e.g., between different L1-varieties or socio-economic groups), as much as such a group can be said to

exist. These debates are beyond the scope of this study (see Granger, 2015). Furthermore, we aim to address several major methodological limitations frequently seen in prior studies, including problems related to corpus comparability and limited sample sizes.

Perhaps the most apparent limitation in previous research on this topic relates to the narrow range of L2 English varieties being investigated. As previously mentioned, by far, the most commonly studied group of L2 English users has been the broadly defined category of L1 Chinese EFL writers. Reasons for this focus include the fact that this group represents a large demographic that frequently travels to English-speaking countries for post-secondary studies. Therefore, it has been deemed essential to understand how this population uses LAs in their target language output.

This focus on L1 Chinese writers has led to a comparative dearth of research on other L1 groups, necessitating further studies to better understand the production characteristics of different EFL communities. The process of expanding the range of L1s has begun in a limited fashion, with research on a small number of additional L1 groups, including Indonesian (e.g., Oktaviant & Sarage, 2022), Arabic (Appel, 2020a; Modhish, 2012), Spanish (Carrio-Pastor, 2013), and Korean (Appel, 2020b; Ha, 2016). Nonetheless, more research is needed, particularly for varieties that have remained under or unstudied.

The second major limitation that our study aims to address pertains to the issue of the native speaker fallacy and the resultant implications of comparisons. Due to the inherent pedagogical implications, scholars have often chosen to highlight differences between L1 and L2 English varieties in hopes of identifying points of deficit in L2 English writing that can then be addressed via targeted teaching interventions. Unfortunately, this deficiency-based view of L2 English writing fails to recognize the value of linguistic production from diverse English-speaking communities, while also overlooking the question of which L1 group should be privileged as the reference ‘native-speaker’ group and the type of writing produced by each group. Issues resulting from this approach include substantial differences in terms of target language proficiency, genre, type, length, and others.

In Chen (2006), an L2 English corpus comprising a wide variety of master’s level texts (e.g., diaries, instructional papers, research proposals) from L1 Chinese EFL writers was compared to a reference corpus of published academic English articles. As the target audience, text type, and length all varied substantially between the two corpora, it is difficult to make claims linking identified production tendencies to any single factor (e.g., L1). This same issue can be found in Ha

(2016) who compared a corpus of L1 Korean EFL writing to writing from the Louvain Corpus of Native English Essays (LOCNESS). While proficiency differences may also have influenced findings in this case, substantial differences in average sample size (L1 English = 855; L1 Korea = 431) created an additional confounding factor. Longer texts may require different organisational approaches, thus impacting the type and token counts for LAs in each corpus. In particular, it is conceivable that longer texts by L1 speakers rely more heavily on a range of discourse markers, including LAs, while there may also be a correlation between shorter texts by L2 writers and a higher frequency of simple sentence structures. This, in turn, may influence the type and frequency of LAs. While these remarks are speculative, the advantages of comparing more similar corpora, in terms of length, type, purpose (knowledge display versus skills display), and proficiency should be clear.

Although a limited number of studies have compared L2 English users in the absence of a native English reference corpus, the aforementioned limitations regarding corpus comparability have largely remained. These shortcomings relate to a lack of adequate controls in the corpora being used, particularly in relation to writing conditions and language proficiency. For instance, studies making use of the International Corpus of Learner English (ICLE), one of the most frequently used data sources for CIA analyses (e.g., Altenberg & Tapper, 1998; Granger & Tyson, 1996), have often grouped all writing from L1 groups together on the basis that ICLE writers have been succinctly described as ‘advanced’ (e.g., Gilquin, 2008). However, it is important to recognize that the writing included in the ICLE is sourced from institutions worldwide, where access to materials, writing conditions, and target language proficiency may vary widely. Therefore, it is extremely difficult to claim that any single factor, such as L1 background, might be the source of production differences. With reference to potential proficiency differences represented in the ICLE, a small-scale analysis suggests proficiency levels vary greatly, with essays found to fall between B2 and C2 on the Common European Framework of Reference (Granger & Thewissen, 2005). As such, results from studies making use of this dataset for CIA analyses should be interpreted with caution.

Finally, the limited sample sizes often used in previous research of this kind deserve mention. This is evident in several studies, each of which has made use of fewer than 20 L2 English writers per group (e.g., Bolton et al., 2002; Carrio-Pastor, 2013; Lei, 2012). Due to these small sample sizes, it remains uncertain how well the results from these studies represent the larger

populations from which they are drawn. Therefore, it is necessary for future studies to expand the number of participants in order to offer a better representation of each target language group.

3. The Current Study

This study aimed to use a more closely controlled and comparable corpora of L2 English academic writing to better identify potential L1-related production tendencies in the use of LAs by writers from three distinct L2 English communities: Filipino, Pakistani, and Thai. While the focus on corpus comparability was a major factor in this research, we also aimed to target understudied L2 English varieties by using data from groups that have only received very limited attention in previous research, particularly in relation to the language feature being investigated in this study (i.e., LAs). The main research question for this study is:

- 1) How does the use of LAs in the L2 English academic writing of university students in the Philippines, Pakistan, and Thailand vary?

4. Methodology

4.1 Corpora

All writing for each of the three targeted L2 English varieties was collected from version 2.3 of The Written Essay Module of the ICNALE. The ICNALE is a highly controlled source of data that can be used for interlanguage contrasts (Ishikawa, 2013). While the corpus itself contains four modules (spoken monologue, spoken dialogue, written essays, edited essays), the written essay portion (unedited) was the sole data source used in this study. Furthermore, to create a more tightly controlled collection of writing, only essays written in response to a single prompt (*It is important for college students to have a part time job*) were included, so as to control for any potential impact a topic may have on the language produced.

Data collection for the ICNALE in each country followed the same general procedure in which students were given 20-40 minutes to respond to an assigned prompt and they composed their texts using a computer. After data gathering, collected essays were assigned proficiency levels based on the Common European Framework of References for languages (CEFR) using a combination of standardized L2 vocabulary size tests and results from previously completed high-stakes English proficiency assessments (e.g., TOEFL, TOEIC).

To control for variations in proficiency levels, only those essays evaluated to be at the B1_2 (B1 high) level were included in the analysis for the current research. As this proficiency level also contained the largest number of essays from most L1 groups represented in the ICNALE, selecting essays from this proficiency level allowed for a larger volume of data in the subsequent analysis. Based on random selection, 80 essays from each target group were collected. The corresponding corpus statistics are provided in Table 1.

Table 1: Corpus Statistics

	Pakistan	Philippines	Thailand
Total Running words	18,700	20,053	18,419
Mean (range)	234 (198 – 331)	251 (200 – 326)	230 (186 – 321)
Standard deviation	33	36	27
Total Sentences	1,136	977	1,039
Mean (range)	14 (6 – 30)	12 (4 – 20)	13 (6 – 22)
Standard Deviation	4	4	4

4.2 Data Extraction

To accurately identify all LAs in each collection of writing, manual analysis based on careful reading of each text was used. This was viewed as necessary, since qualification as an LA is often context dependent, punctuation may be inconsistent, and novel uses of LAs have been identified (e.g., *in another hand*, *final*), particularly among L2 English users. Furthermore, as stated by Bolton et al. (2002), LAs should be viewed as a non-exhaustive list of discourse structuring devices. Thus, automated extraction based on predefined lists may fail to provide an accurate representation of the full range of LAs being used.

A research assistant was hired and trained before the commencement of independent LA extraction. The training for this research assistant focused on two main identifying criteria: semantic function and syntactic independence. Semantic function relates to the role of LAs as discourse structuring devices that are generally found at, or above, the sentence level and should be viewed as providing a semantic link between sections of discourse. In other words, LAs carry the semantic function of indicating how the writer believes the ensuing information should be interpreted in light of what has already been presented. This is in contrast to simply offering a viewpoint that guides the subsequent text (e.g., *in my opinion*). One exception to this criterion was made in the case of time statements, such as *nowadays* (functionally classified as a transition signal), which were found frequently at the beginning of students' writing. Although this item was

commonly used to begin a piece of writing, and thus unable to link previously presented material with the following content, it was included in the present study as it seemed to be an attempt by the writer to link their text to the general situation in the world, thus providing a sort of semantic link.

The second feature, syntactic independence, helps distinguish LAs from other discourse structuring devices (e.g., subordinators, coordinating conjunctions). LAs are relatively unique in terms of the fact that they can be removed from the sentence in which they are found without impacting grammatical acceptability of the utterance. Thus, in contrast to coordinating conjunctions and subordinators, which have both semantic and syntactic functions, LAs are purely semantic in nature.

Subsequent to explaining the extraction criteria to the research assistant and providing representative examples from previous studies, the lead author of this study and the trained research assistant independently reviewed 5 sample essays from 8 different language groups included in the ICNALE. After reconvening to discuss potential discrepancies, it was found that interrater agreement was above 90%, thus, the research assistant was able to proceed with the remaining LA extraction independently. However, the lead author continued to work with the research assistant whenever questionable use was encountered in order to promote consistency throughout the data extraction.

4.3 Data Analysis

The analysis followed a multi-step process. First, all extracted LAs from each corpus were reviewed to provide general findings. Next, to better identify intragroup tendencies and intergroup differences, as well as to avoid assigning idiosyncratic tendencies of a small number of writers to the L1 group to which they belong, raw findings were pruned to focus only on those LAs which appeared in the writing of at least 5 writers from each L1 (i.e., range of 5 texts).

Next, all LAs in each corpus which met the minimum range criterion were assigned a functional category using the taxonomy initially introduced by Quirk et al. (1985). However, following Appel and Szeib (2018), the largely overlapping resultative and inferential categories were combined (simply listed as ‘resultative’). Thus, in total, six major functional categories were included in the present study: listing (see below), summative (e.g., *overall*), appositional (e.g., *for example*), resultative (e.g., *in this way*), contrastive (e.g., *however*), and transitional (e.g., *besides*).

The major functional category of listing devices was further separated into additive (e.g., *moreover, furthermore*) and enumerative (e.g., *first, second*) subcategories to better highlight potential production differences.

Finally, as a way of identifying significant intergroup production differences related to functional categories and individual items, frequency figures were analyzed using log-likelihood statistics by way of Rayson's online calculator¹, as this method has previously been highlighted as a suitable method of identifying intergroup production differences in corpus studies, including those of differing sizes (Rayson & Garside, 2000). Since LAs commonly appear in sentence initial position, general results are listed based on normed occurrences per 1,000 sentences.

5. Results

5.1 General Findings

A total of 829 LAs were identified through the initial manual extraction. This includes 288, 235, and 306 LAs in the corpora from Pakistani, Filipino, and Thai writers, respectively. A clear distinction can be seen, with Pakistani and Thai writers employing a relatively high number of LAs in their writing compared to Filipino writers who made much rarer use of these discourse structuring devices. However, a secondary, potentially more significant, pattern was identified, showing clear differences in production between L1 Pakistani speakers, and L1 Filipino and L1 Thai speakers. In fact, very few statistically significant differences between the latter two groups were identified (see below). As previously mentioned, to avoid attributing idiosyncratic production from a small number of writers to their respective L1 groups, raw figures were pruned to focus only on those LAs appearing in the writing of at least five different users from at least one of the three L1 corpora. These revised findings serve as the focus for all subsequent analyses.

Table 2 presents the top 10 most frequently occurring LAs from each L1 group. While there is substantial overlap with several items being highly frequent in each corpus (e.g., *so, and, but, [now]adays*), unique production tendencies emerged. For instance, Pakistani writers were dependent on a much smaller range of LAs in their writing when compared to Filipino and Thai writers of English. This is evidenced by the fact that Pakistani writers had only 8 items which meet the minimum range threshold (the lowest of any L1 group). Furthermore, these writers were

¹ <https://ucrel.lancs.ac.uk/llwizard.html>

heavily dependent on the top five most frequently occurring items. This is largely due to the high frequency of two particular LAs, *so* and *but*, both of which occur with much greater frequency among Pakistani writers. This discrepancy is evident both in terms of normed frequency and the percentage of total LAs.

Table 2: Top 10 Most Frequently Occurring Linking Adverbials

Pakistan		Philippines		Thailand	
Linking Adverbial	Frequency	Linking Adverbial	Frequency	Linking Adverbial	Frequency
so	59.9 (31%)	but	27.6 (18%)	and	34.6 (15%)
but	53.7 (28%)	now(adays)	19.4 (13%)	but	29.8 (13%)
and	25.5 (13%)	and	16.4 (11%)	so	24.1 (10%)
now(adays)	18.5 (10%)	so	15.4 (10%)	however	21.2 (9%)
in this way	16.7 (9%)	however	13.3 (9%)	now(adays)	18.3 (8%)
also	7.9 (4%)	thus	10.2 (7%)	for example	15.4 (7%)
on the other hand	7.9 (4%)	also	8.2 (6%)	moreover	13.5 (6%)
hence	4.4 (2%)	first (of all)	8.2 (6%)	therefore	11.5 (5%)
		for example	7.2 (5%)	too	10.6 (5%)
		thus	7.2 (5%)	(the) second	9.6 (4%)

Note: Based on normed frequencies per 1,000 sentences

5.2 Functional Analysis

Functional analyses were also carried out on the collected data using a revised version of the taxonomy originally introduced by Quirk et al. (1985). In assessing functional category production differences, normed production figures were compared using Rayson's online log-likelihood calculator. As shown in Table 3, statistically significant differences were identified for overall LA production (tokens), as well as three major functional categories (listing, appositional, resultative) and two subcategories (additive, enumerative).

Table 3: Linking Adverbials by Functional Category

Functional Category	Pakistan	Philippines	Thailand
Listing	33.5 (17%)	45 (30%)	*96.2 (41%)
Additive	33.5 (17%)	24.6 (16%)	*75.1 (32%)
Enumerative	*0 (%)	20.5 (14%)	21.2 (9%)
Summative	0 (%)	0 (%)	0 (%)
Appositional	*0 (0%)	7.2 (5%)	15.4 (7%)
Resultative	*81 (42%)	32.8 (22%)	41.4 (18%)
Contrastive	61.6 (32%)	46.1 (31%)	55.8 (24%)
Transitional	18.5 (10%)	19.4 (13%)	24.1 (10%)
TOTAL	194.5 (100%)	*150.5 (100%)	232.9 (100%)

* $p < .01$

Based on normed frequency counts ($p < .01$), L2 English writers from the Philippines produced the fewest total LA tokens. Although these writers were the least frequent overall users of LAs, no unique production tendencies (significantly low/high frequency) related to any of the major functional categories could be identified. Despite a lack of statistical significance, these writers displayed a clear preference for contrastive (31%) and listing (30%) LAs.

In contrast to Filipino writers of L2 English, and notwithstanding similar overall production tendencies between the two corpora (see below), Thai writers were the most frequent overall users of LAs based on normed counts. This tendency seemed to be primarily related to a high frequency of listing devices, with the highest normed token count (96.2) for any functional category among the three writer groups ($p < .01$), and the highest percentage of overall production (41%) for any functional category. In fact, the normed frequency count for this category among Thai writers was greater than the combined total found in the Pakistani and Filipino corpora. A closer look at this major functional category also indicated that the usage tendency was predominantly attributed to the use of additive LAs, which accounted for 74% of LA tokens in this category. Once again, the total occurrences of items from this functional subcategory among L1 Thai writers surpassed the combined figure from Pakistani and Filipino writers.

Pakistani writers of L2 English were found to hold the middle ground between Thai and Filipino writers in terms of total LA tokens. However, a closer look at the distribution of functional (sub)categories revealed a notable distinction. Pakistani writers' use of listing devices was exclusively confined to the additive subcategory, completely avoiding enumerative LAs. L1 Pakistani writers were found to display statistically significant lower frequency usage of these LAs in their writing when compared to the other groups of L2 English writers.

In addition to the lack of enumerative LAs in their texts, Pakistani writers also did not utilise any appositional devices, a finding that yielded a statistically significant difference compared to other writer groups. In contrast to these low production tendencies, Pakistani writers demonstrated a relatively high frequency in one functional category: resultative. Finally, while all writer groups showed a preference for contrastive and resultative LAs, Pakistani writers were the most dependent on these categories, with nearly 74% of all LA occurrences serving these functions.

5.3 Individual LA Production Differences

For individual item contrasts, all LA occurrences in each corpus were used, regardless of whether frequency counts met the minimum range criteria. For example, although *thus* did not meet the minimum range threshold in both the Pakistani and Thai corpora as a result of having only two occurrences in each collection of writing, these figures (as opposed to 0) were used when running log-likelihood tests. This decision was made as we believe it provides more accurate intergroup comparisons, particularly for low frequency LAs. Just like in the functional category analyses, results for individual LA differences are only highlighted if these contrasts yielded a significance level of $p < .01$.

5.3.1 Pakistani writers of L2 English

For Pakistani writers, unique production tendencies were revealed in a total of 5 individual LAs. These were split between low-frequency usage (*[the] second [of all], however*) and high-frequency usage (*but, so, in this way*). Perhaps unsurprisingly, the high-frequency production of both *so* and *but*, as previously highlighted in Table 2 was found to be statistically significant. These items were not only highly frequent in Pakistani L2 English writing but may well be characteristic of the writing produced by this group. Further research is needed to explore this hypothesis, and to determine whether learner proficiency, individual writing tendencies, and the specific L1 language of the individual writers may also account for these frequency patterns.² In the case of *so*, this item helps explain the statistically significant high frequency of the resultative functional

² A socio-linguistic analysis is beyond the scope of this paper; however, Pakistan is linguistically diverse. While most languages used in Pakistan come from the same linguistic family, the status of languages in Pakistan varies widely and can have a very real impact on the quality of an individual's English education (Mansoor, 2004; Tamim, 2014).

category, as this single item represented 74% of all tokens from this class. However, the frequent use of *in this way* also seems significant, accounting for nearly all of the remaining tokens performing this function (21%). Therefore, over 90% of all resultative LAs used by Pakistani EFL writers can be attributed to these two high-frequency items.

In terms of low-frequency items, one LA each from the enumerative (*[the] second [of all]*), and contrastive (*however*) categories was identified. While the low-frequency usage of *second* does help explain the previously identified statistically significant low frequency of enumerative devices by Pakistani writers of English, it should be noted that of all 4 items included in this category (*first, finally, second, lastly*), only *first* (2 tokens) had any occurrences in the Pakistani corpus. Therefore, the lack of statistically significant low frequency related to individual items from this category should be examined more closely in future studies in order to better understand the general pattern revealed here.

5.3.2 Filipino writers of L2 English

As with the functional analysis, no unique individual production tendencies for the Filipino corpus reached statistical significance. This seemed to be primarily due to a relatively high degree of similarity between the Filipino and Thai corpora, as few production tendencies between these two groups reached the $p < .01$ level of significance. Given the stark difference in total token counts between these two writer groups, this was somewhat surprising (to be discussed in Section 6).

While high frequency (e.g., *second, last, thus*) and low frequency (*so, in this way, but*) usage tendencies were apparent when compared to Pakistani writers, these differences failed to reach significance when Thai writers were included in the comparison. In fact, only two significantly different production tendencies were identified between Filipino and Thai writers (i.e., a higher frequency of *and, in [the] present* among Thai writers). In light of these findings, it is important to highlight the fact that, although the focus of this study was the identification of unique production tendencies that help distinguish each L1 group, results from these contrasts are highly dependent on the choice of comparison group(s).

5.3.3 Thai writers of L2 English

For L1 Thai writers, only one unique production tendency was identified - the high frequency of *in (the) present*. With no occurrences in any of the other corpora, use of this item is highly indicative of Thai EFL writing. However, with only 6 total occurrences, the pattern may not be widespread and future studies should aim to either support or refute this finding.

Similar to the lack of significant findings in the Filipino corpus, the limited number of significant individual item production tendencies in the Thai corpus seems to be related to the high similarity between Filipino and Thai writers. Thus, despite several unique production tendencies identifiable when compared with Pakistani writers (e.g., *furthermore*, *moreover*, *finally*), the inclusion of Filipino writers in the comparison group meant that these production tendencies failed to meet the minimum intergroup significance threshold. Although this study found a lack of significant individual item production tendencies among Thai and Filipino writers, the overlapping nature of these two writer groups is, in itself, a noteworthy finding that warrants further exploration.

6. Discussion

Using a CIA approach, this study examined academic writing from three distinct L2 English varieties to discover unique intragroup production tendencies and intergroup production differences that distinguish each understudied linguistic group. These differences ranged from general production tendencies (tokens and types) to preferred functional categories, and individual items. In the following sections, main findings related to each group of writers are discussed.

6.1 Pakistani writers of English

Pakistani writers of English held the middle ground between Thai and Filipino writers in terms of total LA occurrences (normed token counts). Despite a moderate presence of LAs, these writers displayed the narrowest range of items meeting the range threshold (5 texts), with only 8 total LAs. In other words, these writers exhibited the greatest dependence on a limited range of LAs to structure their writing. This was perhaps most evident through the particularly high frequency of two items, *so* and *but*, which accounted for 59% of total LA occurrences. The high frequency of

these items stands in contrast with the other two writer groups, as the top two items used by Filipino and Thai writers accounted for only 32% and 28%, respectively. The high frequency of these two items also indicates a dependence on coordinating conjunctions as LAs to structure text and make connections between sentences. Since this tendency has been suggested to be more conversational in approach (Poole, 2018), the essays produced by these writers may be viewed as more casual in style. However, it should be noted that all writer groups in this study seemed to favor the use of coordinating conjunctions as sentence-linking devices.

In terms of functional category differences, Pakistani EFL writers in this study were unique due to their absence of any enumerative and appositional devices. These production tendencies may be characteristic of these writers, as support for these findings can also be seen in Jameel et al. (2014) and Mahmood and Ali (2011). For instance, Jameel et al. (2014) found two enumerative (*first, second*) and two appositional (*for instance, for example*) LAs to be relatively infrequent in L1 Pakistani EFL writing when compared to native-English, Japanese, and Chinese writers of English. Similarly, Mahmood and Ali (2011) found a relatively infrequent use of the enumerative LAs *first* and *last* when comparing a corpus of Pakistani English composed of a wide variety of text types (e.g., religious, legal, newspaper, legal) to two corpora of native English varieties (American and British). However, given the lack of information regarding critical methodological details in each of these studies (e.g., corpus size, inclusion criteria), further research is needed to better understand production tendencies related to this population of writers.

Echoing the previous discussion regarding the limited range of LAs, the focus on enumerative and appositional devices by these writers further underscores the limited range of options utilised. Although writers from the Philippines and Thailand only made moderate use of appositional LAs, the lack of enumerative LAs in the Pakistani corpus does contrast quite sharply with the other two corpora. Given that the texts analysed in this study were argumentative in nature, this suggests an argumentative approach that neglects appositional and enumerative relationships in order to place more emphasis on the number of arguments or points being made.

6.2 Filipino writers of English

In terms of total LA production, L2 English writers from the Philippines were found to use LAs the least frequently, according to normed frequency counts. Although this could be interpreted as a

sign of decreased attention to the importance of discourse structuring in Filipino EFL writing, it might place Filipino writers more in line with L1 English norms, as previous research has revealed a relative lack of LAs among L1 English writers when compared to EFL groups (e.g., Appel, 2020a). Thus, these writers may be using alternative methods, such as syntactic overlap or sentence structuring, to achieve coherence in their writing. Further research will be required to confirm or disprove this hypothesis.

Interestingly, despite the relative scarcity of LAs in Filipino EFL writing, no unique functional category production tendencies were identified. As previously stated, this seemed to be largely a result of the similarity between the Filipino and Thai corpora, which was apparent despite a wide discrepancy in total token counts. With a lack of LA focused studies targeting Filipino and Thai writers of English, it is difficult to provide any basis for our findings. The fact that Filipino and Thai derive from unrelated language families, and each country has had a distinct relationship with colonial powers during their respective modernizing periods could be influential factors. This point also bears on each country's respective education system, with significant differences ranging from policy to curriculum to funding (UNESCO, 2014). Interestingly, Perez-Amurao and Sunanta (2020) found that Filipinos have recently become the largest group of English teachers in Thailand by number, a finding that may partially account for the similarity in the two groups' production tendencies. However, these observations are necessarily tentative.

6.3 Thai writers of English

Thai writers were found to be the most frequent users of LAs as a whole, yet this discrepancy did not reach statistical significance. However, a statistically more frequent production was identified in relation to the listing category. A deeper dive into this production pattern revealed that this could be largely attributed to use of additive items, which accounted for 79% of total occurrences. Comparing normed frequencies for this functional subcategory with the other writer groups also indicated a statistically significant difference. Without any identified individual LA production tendencies contributing to these unique production tendencies, it would seem that Thai writers rely broadly on this functional category, rather than possessing a strong focus on a limited number of options.

Previous research offers limited support for the conclusion that additive items tend to be favoured by Thai EFL writers. This can be seen in Jangarun and Lukasaneeyanawin (2016), where

additive linking devices were revealed as the most commonly employed category by both Thai and native-English university writers. Furthermore, previous studies have suggested that *and*, which would fall under the additive subcategory of listing devices is also highly frequent in Thai EFL users' spoken and written discourse (e.g., Arya, 2020; Prommas & Sinwongsuwat, 2011). However, as with earlier studies on Pakistani EFL writing, the lack of methodological details in these studies imply the need for future research to confirm or refute these tentative conclusions.

7. Implications

In terms of methodological implications, the results stress the value of LAs in highlighting L1-related production differences. As a result, more in-depth explorations of usage patterns for these items may prove useful in author attribution research (Jarvis & Crossley, 2012). Additionally, this study highlights the importance of the choice of comparison groups in contrasts of this kind, as the relative lack of unique production tendencies in Thai and Filipino writers seemed to be largely related to the similarity between these two writer groups. Thus, future studies may choose to explore a wider range of linguistic communities to better identify intra/intergroup and shared/unique production tendencies.

In terms of pedagogical implications, the frequent use of coordinating conjunctions functioning as LAs suggests a somewhat conversational tone to the writing from all L2 English user groups. As a result, it may be beneficial for all writer groups to receive more instruction on common academic English writing features, such as the avoidance of coordinating conjunctions in sentence-initial position. That being said, depending on the field of study, this usage pattern has varying degrees of acceptance in academia (Poole, 2018). Nevertheless, the strong dependence of specific writer groups on a narrow set of LAs should be addressed pedagogically. This was perhaps most apparent in the Pakistani corpus, where only eight LAs met the frequency threshold and 72% of total LA use was related to three items (*so*, *but*, *and*). Since these items are also coordinating conjunctions, targeted instruction on how to avoid these items in favor of more academic choices (e.g., *thus*, *however*, *furthermore*) would likely prove beneficial, particularly in the early stages of students' academic English education when simple rules may be preferable.

Given the unique production tendencies displayed by all L1 groups, it would be beneficial to highlight these patterns to each writer group. This would help them in recognising and adjusting their tendencies accordingly. Given the freely accessible nature of the ICNALE, in class activities

in which students perform searches for common LAs in the collected texts produced by writers from their linguistic community in comparison to other linguistic communities could help highlight these patterns so that students can begin to move beyond their existing tendencies.

8. Limitations and Future Research

There are several important limitations that should be addressed in future research. First, despite efforts to control for proficiency level differences, the method used by the ICNALE design team to assign proficiency scores may have some shortcomings. They primarily rely on previous scores from standardized language test and vocabulary size assessments, which did not assess the actual language produced by each writer. Therefore, future research could benefit from actual proficiency assessments of the language being analysed. In addition, as the Written Essay Module of the ICNALE is exclusively composed of argumentative essays, it would be beneficial to expand the range of writing styles examined, ideally within the same study, to better understand how these factors impact LA use.

Additionally, due to the exploratory nature of this research, the focus resided solely in the discovery of unique production tendencies that could be attributed to each L1 group and not the underlying basis of each tendency. Future research should aim to identify potential root causes of the production patterns highlighted in this study. To do so, it will be necessary to examine potential L1 translation equivalents, analyse L1 corpora from each target variety, review English teaching materials used in the respective countries, and evaluate the general teaching approach most commonly applied during instruction.

9. Conclusion

This study used a CIA approach to examine the usage patterns of LAs in the L2 English academic writing of post-secondary students from three countries: Pakistan, Philippines, and Thailand. Through both quantitative and qualitative analyses, unique individual item production tendencies, functional category patterns and general production tendencies were identified. For instance, Pakistani writers displayed the greatest dependence on a limited range of LAs, Filipino writers showed less frequent use of LAs, aligning more with L1 English norms, while Thai writers exhibited frequent use of additive items. These findings contribute to the growing body of research

into L2 English production tendencies and suggest that EFL writers from different linguistic backgrounds tend to structure their texts in diverse ways.

References

- Appel, R. (2020a). A contrastive Interlanguage analysis of linking adverbials in EFL writing: Identifying L1 related differences. *Learner Corpus Studies in Asia and the World*, 4, 24-40. <https://doi.org/10.24546/81011992>
- Appel, R. (2020b). An exploratory analysis of linking adverbials in post-secondary texts from L1 Arabic, Chinese and English writers. *Ampersand*, 7, 100070. <https://doi.org/10.1016/j.amper.2020.100070>
- 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, 35-58. <https://doi.org/10.1016/j.jeap.2020.100873>
- Appel, R. & Szeib, A. (2018). Linking adverbials in L2 English academic writing: L1-related differences. *System*, 78, 115-129. <https://doi.org/10.1016/j.system.2018.08.008>
- Altenberg, B., & Tapper, M. (1998). The use of adverbial connectors in advanced Swedish learners' written English. In S. Granger (Ed.) *Learner English on computer* (pp. 3-18). Longman. Retrieved from <https://api.semanticscholar.org/CorpusID:60561736>
- Arya, T. (2020). Exploring discourse marker use in Thai university students' conversations. *LEARN Journal: Language Education and Acquisition Research Network*, 13, 247-267. Retrieved from <https://so04.tci-thaijo.org/index.php/LEARN/article/view/237854>
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. (1999). *Longman Grammar of spoken and written English*. Longman. Retrieved from <https://www.degruyter.com/database/COGBIB/entry/cogbib.1245/html>
- Bolton, K., Nelson, G., Hung, J. (2003). A corpus-based study of connectors in student writing: research from the international corpus of English in Hong Kong (ICE-HK). *International Journal of Corpus Linguistics*, 7, 165–182. <https://doi.org/10.1075/ijcl.7.2.02bol>

- Carrio-Pastor, M. (2013). A contrastive study of the variation of sentence connectors in academic English. *Journal of English for Academic Purposes*, 12, 192–202. <https://doi.org/10.1016/j.jeap.2013.04.002>
- Chapwanya, F., & Nel, J. (2023). Discourse markers *so* and *well* in Zimbabwean English: A corpus-based comparative analysis. *World Englishes*, 00, 1-24. <https://doi.org/10.1111/weng.12626>
- Chen, C. (2006). The use of conjunctive adverbials in the academic papers of advanced Taiwanese EFL learners. *International Journal of Corpus Linguistics*, 11, 113–130. <https://doi.org/10.1075/ijcl.11.1.05che>
- Crewe, W.J. (1990). The illogic of logical connectives. *The ELT Journal*, 44, 316-325. <https://doi.org/10.1093/elt/44.4.316>
- Faya-Cerqueiro, F., & Macho-Harrison, A. (2022). Use of linking words in argumentative written texts: Comparison between L1 and L2. *Revista de Estudios Sobre Lectura*, 21, 1-19. https://doi.org/10.18239/ocnos_2022.21.2.2884
- Gilquin, G. (2008). Combining contrastive and interlanguage analysis to apprehend transfer: detection, explanation, evaluation. In G. Gilquin, S. Papp, M.B. Diez-Demar (Eds.), *Linking up Contrastive and Learner Corpus Research* (pp. 3-33). Rodopi. https://doi.org/10.1163/9789401206204_002
- Granger, S. (1998). The computer corpus: A versatile new source of data for SLA research. In S. Granger (Ed.) *Learner English on computer* (pp. 3-18). Longman.
- Granger, S. (2015). Contrastive interlanguage analysis: a reappraisal. *International Journal of Learner Corpus Research*, 1, 7–24. <https://doi.org/10.1075/ijlcr.1.1.01gra>
- Granger, S., Thewissen, J. (2005, June). Towards a reconciliation of “Can Do” and “Can’t Do” approach to language assessment. June. In *Paper Presented at the Second Annual Conference of the European Association of Language Testing and Assessment*, Voss, Norway.
- Granger, S., Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, 15, 17–27. <https://doi.org/10.1111/j.1467-971X.1996.tb00089.x>

- Ha, M.-J. (2016). LAs in first-year Korean university EFL learners' writing: a corpus informed analysis. *Computer Assisted Language Learning*, 29, 1090–1101. <http://dx.doi.org/10.1080/09588221.2015.1068814>
- Hsu, C., Davis, R., & Wang, Y-C. (2020). Chinese learners' use of concessive connectors in English argumentative writing. *Concentric*, 46, 95-123. <https://doi.org/10.1075/consl.00014.hsu>
- Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In: Ishikawa, S. (Ed.), *Learner Corpus Studies in Asia and the World, vol. 1*. Kobe University, Kobe, pp. 91–118. Japan. <https://doi.org/10.24546/81006678>
- Jameel, I., Mahmood, M., Hussain, Z., Shakir, A. (2014). A corpus-based analysis of linking adverbials in Pakistani English. *International Journal of Linguistics*, 6, 133-140. <https://doi.org/10.5296/ijl.v6i3.5332>
- Jangarun, K., & Luksaneeyanawin, S. (2016). Discourse connector usage in argumentative essays by American and Thai university students. *Journal of Pan-Pacific Association of Applied Linguistics*, 20, 95-112. Retrieved from <https://eric.ed.gov/?id=EJ1110802>
- Lei, L. (2012). Linking adverbials in academic writing on applied linguistics by Chinese doctoral students. *Journal of English for Academic Purposes*, 11, 267–275. <https://doi.org/10.1016/j.jeap.2012.05.003>
- Liu, D. (2008). LAs: An across-register corpus study and its implications. *International Journal of Corpus Linguistics*, 13, 491–518. <https://doi.org/10.1075/ijcl.13.4.05liu>
- Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, 29, 16-27. <http://dx.doi.org/10.1016/j.jslw.2015.06.003>
- Mahmood, M., & Ali, N. (2011). Circumstance adverbials in Pakistani English a corpus-based comparative analysis. *International Journal of Academic Research*, 3, 47-51.
- Mansoor, S. (2004). The status and role of regional languages in higher education in Pakistan. *Journal of Multilingual and Multicultural Development*, 25(4), 333-353. <http://dx.doi.org/10.1080/01434630408666536>
- Milton, J., & Tsang, E. S. C. (1993). A corpus-based study of logical connectors in EFL students' writing: Directions for future research. In R. Pemberton & E.S.C. Tsang (Eds.) *Studies in*

- lexis* (pp. 215-246). The Hong Kong University of Science and Technology Language Centre. Retrieved from <https://hdl.handle.net/1783.1/1083>
- Modhish, A. (2012). Use of discourse markers in the composition writings of Arab EFL learners. *English Language Teaching*, 5, 56-64. <http://dx.doi.org/10.5539/elt.v5n5p56>
- Oktavianti, I., & Sarage, J. (2022). Linking adverbials in Indonesian EFL students' essays: A corpus-driven study. *Journal on English Language Teaching & Learning, Linguistics and Literature*, 10, 1583-1600. <https://doi.org/10.24256/ideas.v10i2.3061>
- Perez-Amurao, A. L. & Sunanta S. (2020). They Are 'Asians Just Like Us': Filipino Teachers, Colonial Aesthetics and English Language Education in Thailand. *Sojourn: Journal of Social Issues in Southeast Asia*, 35(1), 108–137. DOI:10.1355/sj35-1d
- Poole, R. (2018). *A guide to using corpora for English language learners*. Longman. <https://doi.org/10.1515/9781474427180>
- Prommas, P., & Sinwongsuwat, K. (2013). A comparative study of discourse connectors used in argumentative compositions of Thai EFL learners and English-native speakers. *Tennessee Foreign Language Teaching Association Journal*, 4, 88-102. Retrieved from <https://api.semanticscholar.org/CorpusID:110206044>
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.
- Rayson, P. & Garside, R. (2000). *Comparing corpora using frequency profiling*. *Proceedings of the workshop on comparing corpora*, 9, 1-6. <https://doi.org/10.3115/1117729.1117730>
- Shin, Y. (2019). Do native writers always have a head start over nonnative writers? The use of lexical bundles in college students' essays. *Journal of English for Academic Purposes*, 40, 1-14. <https://doi.org/10.1016/j.jeap.2019.04.004>
- Tamim, T. (2014). The politics of languages in education: Issues of access, social participation and inequality in the multilingual context of Pakistan. *British Educational Research Journal*, 40(2), 280-299. <https://doi.org/10.1002/berj.3041>
- United Nations Educational, Scientific and Cultural Organization. (2014). Education systems in ASEAN+6 countries: A comparative analysis of selected educational issues. Thailand: UNESCO Bangkok. Retrieved from <https://www.right-to-education.org/resource/education-systems-asean6-countries-comparative-analysis-selected-educational-issues>

- Wang, Y. (2022). Linking adverbials in argumentative essays of advanced EFL learners: A corpus-assisted analysis. *Korean Journal of English Language and Linguistics*, 22, 396-417. <https://doi.org/10.15738/kjell.22..202204.396>
- Yeung, L. (2009). Use and misuse of 'besides': A corpus study comparing native speakers' and learners' English. *System*, 37, 330-342. <https://doi.org/10.1016/j.system.2008.11.007>