University of Memphis

University of Memphis Digital Commons

2020

A contrastive corpus analysis on the use of connectors in students writing from 10 Asian countries as compared to native experts: Research from the ICNALE (The International Corpus Network of Asian Learners of English)

Hyun Soon Cho Min

Follow this and additional works at: https://digitalcommons.memphis.edu/etd

A CONTRASTIVE CORPUS ANALYSIS ON THE USE OF CONNECTORS IN
STUDENTS' WRITING FROM 10 ASIAN COUNTRIES AS COMPARED TO
NATIVE EXPERTS: RESEARCH FROM THE ICNALE (THE INTERNATIONAL
CORPUS NETWORK OF ASIAN LEARNERS OF ENGLISH)

by

Hyun Soon Cho Min

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: English

The University of Memphis

August 2020

Acknowledgement

I would like to express my gratitude and appreciation to Dr Teresa Dalle, my committee chair, for her dedicated support and guidance. Dr Dalle continuously provided the thoughtful comments and was always willing and enthusiastic to assist me in any possible way throughout this project. Without her guidance and dedication, this dissertation would not have been possible.

In addition, I would also like to thank my committee members, Dr Emily Thrush, Dr Romy Ghanem and Dr Sage Graham for constructive comments. Especially I am thankful to Dr Ghanem for providing valuable advice and special time regarding statistical analysis.

Finally, my biggest thanks to my family for all the support you have shown me during the compilation of this dissertation. I wish to thank my supportive husband, George, and my two wonderful children, Brandon and Ellie, who are my source of unending inspiration.

Abstract

Hyun Soon (Julie) Cho-Min. PhD. The University of Memphis. August/2020. The contrastive corpus analysis on the use of connectors in students' writing from 10 Asian countries as compared to native experts: Research from the ICNALE (The International Corpus Network of Asian Learners of English)

Major Professor: Dr Teresa S. Dalle

My dissertation focuses on the connector use in the writings of the Asian students from 10 different countries in comparison with that of the English native speakers. Specifically, I examine how the Asian students use the connectors similarly or differently compared to English speakers in terms of frequency and choice, depending on their respective countries and English proficiency levels.

The research questions addressed in this study are these: Is there a similarity in the use of connectors between the writing of college students in 10 Asian countries and native speakers (NS) of English, in terms of frequency and choice of connector? If there is, how do they employ the connectors similarly? Is there a difference in the use of connectors between the writing of college students in 10 Asian countries and native speakers of English in terms of frequency and choice of connectors? If so, how do they employ the connectors differently? Is there a difference in the use of connectors between the writings of college students within 10 Asian countries depending their English proficiency level and their nationalities?

In responding these questions, I use the written essay module of the ICNALE (the International Corpus Network of Asian Learners of English), which is available to the public. Two computational tools (Coh-Metrix and AntConc) are used to analyze and identify the commonality and difference on connector usage among the different language groups.

The current study reveals the similarities and differences with which the Asian students and the English speakers use the connectors. One major finding suggests that the Asian students underuse additive and negative connectors compared to their English counterparts and that they prefer to position certain connectors at sentence-initial positions.

By signposting the Asian students' writing patterns as compared to the norm of the English speakers, my dissertation aims to heighten the awareness of connector use and offer some helpful information to language learners and present an important instructional resource to ESL educators and textbook designers about the authentic use of English.

Table of Contents

Table of Contents (continued)

List of Tables

List of Figures

Chapter 1: Introduction

The ability to write well is important to college students in the United States for its significant role in academic success and beyond. It is a vital skill in achieving a satisfactory grade in class because a significant portion of coursework in U.S. colleges demands some form of writing from extensive texts like research papers or reports to even short correspondence with professors and their native colleagues. Moreover, a strong writing ability aids efficient communication with professors and classmates on collaborative projects.

Despite the importance of an effective writing ability, many students experience a challenge in organizing and presenting their thoughts persuasively in adequate academic prose. Its burden is particularly greater to non-native-speaking (NNS) students of English because they need to acquire specific discourse characteristics that are quite different from their native language as well as a general linguistic knowledge of English (Ferris & Hedgcock, 2005; Hinkel, 2002). As a result, it is no surprise to find that NNS students often feel writing in English is the most challenging task to execute (Reid, 1992).

One of the important elements in creating a text is the discourse connector. It is a sub-category of cohesive devices, the words that show the relations between parts of text by referring to previously mentioned information as in pronouns, by repeating the same or similar lexical items, or by using linking words such as "and," "but", etc. The discourse connector is also known as several alternative terms by different researchers such as "conjunctive cohesive devices" (Halliday & Hasan (1976)', "linking adverbials" (Biber, Johansson, Leech, Conrad, and Finegan, 1999), "logical connectors" (Quirk, Greenbaum, Leech, & Svartvik, 1985), "discourse connectors (Cowan, 2008) and discourse connectives" (Blakemore, 2002).

These linking devices are crucial in writing since they combine separate units such as

clauses, sentences and paragraphs, effectively weaving them into unified, well-formed and coherent text (Halliday & Hassan, 1976; Quirk, Greenbaum, Leech, & Svartvik, 1985). Halliday and Hassan (1776) defines these devices:

> Conjunctive elements are not cohesive in themselves, but indirectly, by virtue of their specific meanings; they are not primarily devices for reaching out into the preceding or following text, but they express certain meanings which pre-suppose the presence of other components in the discourse. (p. 226)

By expressing the relationship of ideas, the connectors create interdependency between elements in a text and build connections with one another. Without appropriate use of connectors, a text would not seem logically structured and the relationships between the different units in a text would not be clear. Consequently, it is important to gain a mastery of discourse connectors in order to compose a cohesive writing.

Statement of Problem

For its primary role in connecting different parts in text, the use of the linking devices such as connectors in the production of quality writing has been explored in a large volume of literature for many decades. Especially, many ESL researchers and educators have investigated the use of the discourse connectors in relation to writing quality. Liu and Braine (2005) examined 50 argumentative essays written by Chinese undergraduate students and found there is a significant relation between the number of the discourse connectors and the quality of writing. Similarly, Chanyoo (2018) found a significant relationship on the use of the discourse connectors in 30 academic essays written by Thai undergraduate students and their writing scored rated by the experts. Based on this finding, it has become apparent that the appropriate use of the discourse connector has a share in writing quality.

While the discourse connector plays a positive role in writing quality, a proper use of

the discourse connector has been found a challenge for ESL/EFL learners (Al-Badi, 2005). Moreover, ESL researchers and educators have showed a concern on the language learners' different use of the connectors as compared to English speakers (Crewe, 1990; Hinkel, 2004).

In addressing ESL/EFL learners' problematic area on the discourse connector, Narita, Sato, and Sugiura (2004) compared Japanese sub-corpora of the ICLE (International Corpus of Learner English) to NS corpora from the Louvain Corpus of Native Essay Writing (LOCNESS) on the use of discourse connectors. The result shows the Japanese students overuse certain connectors such as *for example*, *of course*, and *first*, while they significantly underuse such connectors as *then*, *yet*, and *instead*.

In sum, there is a considerable evidence that ESL/EFL learners face challenges in effectively using the discourse connectors. To help facilitate their problematic areas, there are research needs to further identify the use of the discourse connectors that distinguish the writing of ESL/EFL from the writing of English speakers. Specifically, there is a need to consider diverse factors in analyzing ESL/EFL's writings such as linguistic and English proficiency levels. Consequently, the current study seeks to address such a need by isolating and identifying the linguistic difference between ESL/EFL learners and English speakers while considering diverse factors such as their linguistic backgrounds and English proficiency levels.

Purpose of the Study

The current study aims to shed light on the distinctive features of the writings of Asian students in ten countries and compare their use of sentence connectors depending on their level of English and their nationality. In addition, I would like to compare the results with the English speakers' writing and find any similarities and differences in the use of sentence connectors.

Relevance of the study

Over the past decades, of the number of the international students, Asian students have predominated in international student enrollment in the U.S. colleges and universities. According to the U.S. government, 77 percent international students are originated from Asia in the 2016 to 2017 academic year (Student and Exchange Visitor Program, 2017). While they have proved their English proficiency with TOEFL scores upon admittance, most of them express difficulty when they start the classes since most college work in the U.S. is in written form (e.g., reports, essay, summaries, exams, powerpoint slides, even email correspondence). Therefore, there is a need to understand the writing pattern of Asian students so as to offer helpful advice.

Significance of the study

The findings of this study will give an insight into understanding the writing pattern of language learners' use of connectors compared to that of native speakers. The results of this study can be used to provide helpful information to language learners themselves, so that they will have the opportunity to assess their writing habits and, at the same time, English composition writers will receive important data about the authentic use of English from the text of both native speakers and learners.

Research Questions

Previous research shows that Asian students use sentence connectors differently from native speakers. For example, Bolton, K., Nelson, G. & Hung, J (2002) discovered that Hong Kong college students frequently employ certain connectors which differ from those used native speakers. Based on their findings, I formed the following research questions:

(1) Is there a similarity in the use of connectors between the writing of college students in 10 Asian countries and native speakers (NS) of English, in terms of frequency and

choice of connector? If there is, how do they employ the connectors similarly?

(2) Is there a difference in the use of connectors between the writing of college students in 10 Asian countries and native speakers of English in terms of frequency and choice of connectors? If so, how do they employ the connectors differently?

(3) Is there a difference in the use of connectors between the writings of college students within 10 Asian countries depending their English proficiency level and their nationalities? If so, how do they employ the connectors similarly and differently across English level and nationality?

Limitation of the study

While the current study includes a wide list of discourse connectors, it doesn't provide a complete picture of NNS students' use of the discourse connectors. Coh-Metrix generates the occurrence of connectors regardless of their position within a sentence; however, it doesn't provide context where the connectors are used in a text. On the other hand, AntConc offers an immediate context neighboring the connectors so that the researchers can determine their functions within a sentence.

In the initial analysis using AntConc, I include the connectors only at the beginning of sentences, excluding the use of connectors in other sentence positions such as the middle or end. This selective process is necessary because of the large volume of data in this research.

Many of the discourse connectors have different usages. For example, *so* is used as a discourse connector in a sentence like "…*bring benefits in many aspects. So nowadays an increasing number of students are...*" However, *so* is used as an idiomatic expression in a sentence like "… *playing computer games, sleeping and so on.*" To find *so* as the discourse connector, one has to go through each instance and eliminate the other usages. The process can be done when working on small corpus. However, it is impossible to scrutinize each

instance on the current study since it uses a large amount of data. Therefore, in the initial AntConc analysis, I only include the instances when the connectors are in initial position and examine them to ensure they are used as the discourse connectors. However, subsequent analysis can be conducted to examine their different position in a sentence with a smaller set of connectors if it is necessary.

Organization of the study

The next chapter will review a number of researches that report NNS students' difficulty in presenting their thoughts in English, especially in U.S. universities. In addition, the chapter will highlight the linguistic features of writing proficiency and further discuss the studies that examine NNS students' use of cohesive devices and their relation to writing quality. Lastly, the chapter will end with research that focuses on one cohesive device, a discourse connector. The third chapter describes the methodology, data background and collection process for the current study. It also explains the tools, as well as Coh-Metrix and AntConc. The fourth chapter will report on the results of the current study. The final chapter will discuss the results, their implications, and possible future studies.

Chapter 2: Literature Review

NNS students' academic writing difficulties

The ability to write well is important to college students in the United States for its significant role in academic success and beyond. It is a vital skill in achieving a satisfactory grade in class because a significant portion of coursework in U.S. colleges demands some form of writing in the form of extensive texts like research papers, reports, and even short correspondence with professors and their native colleagues. Moreover, a strong writing ability aids efficient communication with professors and classmates on collaborative projects.

Despite the importance of an effective writing ability, many students experience a challenge in organizing and presenting their thoughts persuasively in adequate academic prose. Its burden is particularly greater to second language (L2) students of English because they need to acquire specific discourse characteristics that are quite different from their native language as well as a general linguistic knowledge of English (Ferris & Hedgcock, 2005; Hinkel, 2002). As a result, it is no surprise to find that L2 students often feel writing in English is the most challenging task to execute (Reid, 1992).

Several studies have looked into difficulties faced by international students in their writing of English at the college level. For example, Chou (2011) explored Taiwanese doctoral students' perception of English academic writing. She examined the syllabi of 67 classes and conducted a semi-structured interview with a total of 13 Taiwanese students of differing majors at a university in New York state. The syllabi analysis revealed that all major assignments include some form of writing from online discussion to research proposal. In the interview, Taiwanese students all perceived writing to be extremely important and indeed a major stressor. However, their responses varied across their respective disciplines. Students in humanities and social sciences had a comparatively large quantity of writing assignments

with varying genres of writing. They expressed that the difficulties arise from ambiguous writing instructions from professors, influences from their first language, inaccuracy of grammar use, and lack of content. In contrast, science and technology students had a low number of writing assignments. While they were sure that they could complete class writing assignments, they were still concerned with their grammatical problems and felt they were at a disadvantage when compared with native English speakers in the class.

Similar results were also obtained from Al-Badi's study (2015). He interviewed 20 students of four nationalities (Korean, Chinese, Taiwanese, and Omani) studying at a university in Australia. Seventeen of the students perceived class writing assignments to be difficult. In looking further at their response, the author found that the students ascribed their difficulty to the lack of knowledge of language, coherence, and cohesion. Other factors included difficulty in expressing their own voice, significant topics and relevant references, as well as paraphrasing and referencing a citation correctly. Accordingly, Al-Badi suggests that linguistic knowledge of academic writings in English such as cohesive devices and conventions on academic writings would help NNS students to overcome their perceived problems.

In supporting the NNS students' claims, the professors also showed concern for the overall writing quality of NNS students. Casanave and Hubbard (1992) conducted a survey of 85 professors across multiple fields in a university, regarding their first-year doctoral student writings. More specifically, they explored the features of writing that influenced the grades the professors gave to their students on written assignments. The features were correctness of punctuation/spelling, accuracy of grammar, appropriateness of grammar, size of vocabulary, appropriateness of vocabulary, quality of paragraph organization, quality of overall organization, quality of content, development of ideas, overall writing ability, adequate

treatment of topic, adoption of the appropriate tone/style, and the ability to meet assignment requirements. The findings show that the professors ranked highly the importance of discourse-level criteria (e.g., quality of content, development of ideas, and adequate treatment of topic).

Not surprisingly, all the professors stated that NNS students had more problems than their native colleagues. Regarding the question of the problematic area of NNS students' doctoral writing, the professors highlighted the accuracy and appropriateness of grammar. However, they perceived that NNS students have only minor or moderate problems in meeting the requirements of assignments (e.g., addressing the topic adequately, achieving appropriate tone/style, and meeting the requirements of the assignment).

The findings of the study suggest that the class assignments at a university in the U.S. indeed require significant written work. Both professors and NNS students felt that NNS students' writing was more problematic than their native colleagues. Moreover, they perceived the lack of linguistic knowledge of English as the most problematic area in NNS writing.

Linguistic features of writing performance

Writing difficulty at the college or postsecondary level is not only a concern for NNS students but also for native speakers. To help the students overcome such difficulties, many researchers attempt to offer advice by examining the linguistic features that affect overall writing performance. In this section, I will present several studies that investigate certain linguistic features that distinguish more proficient from less proficient writing. In particular, I will focus on internal characteristics that distinguish higher-rated writing from lower-rated writing among NS writers and compare NS and NNS writers as well.

Cumming, et al (2005) examined 216 compositions written by 36 examinees of three

separate levels of proficiency; the tasks were of varying types. The purpose of the study was to discover the writing quality among integrated tasks (involving writing in response to print or audio source texts) and independent tasks across multiple English proficiency levels of NNS students. Examinees' English proficiency levels were rated as 3, 4 and 5 with 5 being the highest. In examining their discourse differences between task types, the authors also analyzed the linguistic features of writing with higher scores, as compared to those with lower scores.

To determine a varying written performance, Cumming and his colleagues used seven indicators as a guideline in discourse analysis: text length, lexical sophistication, syntactic complexity, grammatical accuracy, quality of argument structure, orientations to source evidence, and verbatim strings of words. Expert raters coded each indicator and their reliability was tested.

Cumming, et al (2005) aimed to find the difference of writing quality across English proficiency levels and task types. I will only present the findings from the comparison across English proficiency levels, which is relevant to my research. First, text length was rated by the total number of words. Lexical sophistication was analyzed in two ways: average word length and type/token ratio of the number of individual lexical words over the total number of words per text. Syntactic complexity was also assessed in two ways: number of clauses and words per each independent clause with all of its dependent clauses. Grammatical accuracy was rated holistically. Quality of argument structure was evaluated by the claims, data, warrants, propositions, oppositions, and responses to oppositions. Orientation to source evidence was coded for presentation of voice in each independent clause together with all of its dependent clauses. Lastly, verbatim strings of words were measured by the number of strings of words in source (reading and listening prompts) that appeared in their produced

writing.

Cummings and his colleagues (2015) have found that writing across multiple English proficiency levels shows differences in all indicators of writing quality. In the indicator of text length, writing between English proficiency 3 and 4, and between levels 3 and 5 has significant differences, but difference was not significant between English proficiency level 4 and 5. The results were consistent in terms of the number of words per composition. As the English proficiency level increased, the number of words per composition also increased, but only between level 3 and 4. That is, the number of words is a distinctive feature between elementary and intermediate level. However, it is not a discernable feature beyond the intermediate level.

Lexical complexity was measured by two indicators: average word length and type-token ratio of the number of separate lexical words over the total number of words per composition. The average word length was relatively consistent across the English proficiency levels. However, the type-token ratio shows significant differences between levels 3 and 4 and between levels 3 and 5. That is, the examinees at level 4 and 5 tended to write more diverse words than those in level 3.

Cummings and his colleagues (2015) measured syntactic complexity in two ways: the number of words and the number of clauses per T-unit. Among the three proficiency levels, examinees showed statistically significant differences. The higher-level examinees used more words per independent clause including all dependent clauses. However, there was no difference in the number of clauses per T-unit. That is, examinees used a similar number of clauses per T-unit regardless of their English proficiency level. Holistic ratings of grammatical accuracy showed significant differences between proficiency levels. As the level increased, the mean ratings of grammatical accuracy also increased.

For the argument's structure, Cummings et al (2015) rated separately the quality of the prepositions, claims, data, warrants, oppositions, and responses to oppositions shown in the examinees' writings. The ratings were significant in all indicators except the quality of claims in arguments, which differed significantly between proficiency levels 3 and 4, and between levels 3 and 5.

In summary, Cummings et al.'s study highlights the features of writing quality among three levels of NNS English proficiency. They found that higher level writing has more words per independent clause when including its dependent clause, and it uses more grammatically correct sentences. However, the number of words in a text shows no difference between intermediate and advanced level, while they are different from those at the elementary level. Additionally, intermediate to advanced level NNS writers use more diverse words than those at the elementary level. The two groups also show consistency in receiving similar ratings for quality of claims. The study also found that the number of clauses in each T-unit and also word length do not distinguish writing performance.

In a similar study, McNamara, Crossley, and McCarthy (2012) explored the linguistic features that distinguish writing proficiency levels among English-speaking writers. They examined a total of 120 essays written by undergraduate students who were enrolled in a freshman English composition class of a U.S. university. The students were given four prompts and were allowed to choose the topic. So there is an unequal number of texts per prompt. Writing was not timed and was done outside of classroom.

In examining the textual characteristics of writing, McNamara et al. (2010) used the automated computation tool, Coh-Metrix, as well as trained raters. Coh-Metrix is an automated text analysis tool developed at the Institute for Intelligent Systems at the University of Memphis. When the user enters an English text, it returns with more than 600

linguistic measures of cohesion, language, and readability (Crossley & McNamara, 2010). Its use on second language research has been validated by several studies (Crossley, Salsbury, & McNamara, 2009).

In the study of McNamara et al. (2010), the raters first graded the essays from 1 to 5 according to a standardized rubric. The essays rated from 1 to 3 were labeled as low-proficiency and the essays at 4 and 5 were the high-proficiency group. Finally, the computation tool, Coh-Metrix, examined the linguistic features in the essays of both the low and high proficiency groups.

The essay quality in both low and high proficiency groups was measured by cohesion (i.e., coreference and connectives), syntactic complexity (e.g., number of words before the main verb, sentence structure overlap), the diversity of words used by the writer, and characteristics of words (e.g., frequency, concreteness, imageability).

The findings show that the more highly-graded essays contained linguistic characteristics related to text difficulty and complicated language. The best indicators of good writing included high syntactic complexity (as measured by number of words before the main verb), lexical diversity, and less frequently used words. That is, skilled writers use more less-familiar words as well as complicated and diverse sentences. However, the study finds the measure of cohesion did not differ between low and high proficient writings.

In another study on writing quality, Taguchi, Crawford and Wetzel (2013) analyzed a corpus of English essays written by NNS students. All of the students were freshmen in a U.S. university and their native languages included Korean, Hindi, Chinese, Thai, Spanish, Russian, German, and French. The essay was a placement test to determine whether the students needed to take ESL composition classes or not as they entered the university. The test was performed online and the students were asked to choose one of two topics and then

read two texts presenting contrasting opinions on the topic. After that, they were asked to write an 850 to 1300-word essay that compared two differing views and which presented their own argument.

Taguchi et al. (2013) compiled a total of 116 essays which were rated by three native speakers of English who had an experience of teaching composition classes at the university. Raters evaluated the essays by five criteria: language use, content, organization, vocabulary, and mechanics. The essays with a score of 90 or above were grouped as high-rated essays while those with below 90 were grouped as low-rated.

When the essays were grouped, Taguch and his colleagues examined them in the context of two categories: *language use* and *content*. The measure of language use was complexity of construction at both clause and phrase level. Specifically, the clause-level complexity was measured by the use of subordinating conjunctions, verb complements, noun complements, adjective complements, that-relative clauses, and wh-relative clauses. The phrase-level complexity included the use of pre-qualifiers, pre-quantifiers, post-determiners, demonstrative determiners, singular definite and indefinite articles, singular or plural determiners, double conjunctions, attributive adjectives, and post-noun-modifying prepositional phrases. The category of content was judged by the degree to which the writers facilitated the source text. It was evaluated by: (1) accurate understanding of, and clear responses to, the source text, and (2) effective use of the source text by direct reference to the authors (i.e., use of author names) and use of specific nouns or verbs that refer to the source text (i.e., advocate, argue, argument, based on, etc.).

The findings on the use of sentence complexity and content reveal that low-rated essays used slightly more clause-level complexity than high-rated essays. However, the in-depth analysis showed differences in the use of clause-level complexity. The low-rated essays

contained more subordinating conjunctions and *that*-relative clauses, while higher-rated essays had more *that*-clause verb complements.

The detailed analysis of phrase-level complexity also shows a difference between lower- and higher-rated essays. Both attributive adjectives and post-noun-modifying prepositional phrases were more frequently discovered in the high-rated essays than the low-rated essays. However, there was no difference in the other measures of phrase-level complexity such as pre-qualifiers, pre-quantifiers, post-determiners, demonstrative determiners, and double conjunctions. In the comparison of essay content, the high-rated essays referred to author names, and they attributed words much more frequently than the lower-rated essays.

NNS's use of cohesive devices

In this section of the literature review, I will briefly discuss the concept of cohesion and then present studies that discussed the use of cohesive devices in NNS writing. The discussion of cohesive devices is pertinent here because my study includes an analysis of one of the cohesive devices.

According to Halliday and Hasan (1976), cohesion is what makes a text a text—a semantic unit of language. It signals the meaningful relation between one part of a text and another part and expresses continuity. Thus, cohesion plays a crucial role in enhancing the reader's understanding of the text (Crossley & McNamara, 2009; Yang & Sun, 2011). In other words, readers can easily understand the text if the writer uses cohesive devices frequently and skillfully.

For its critical role in text comprehensibility, a number of ESL researchers examined the use of cohesive devices in NNS writing, and some further explored its relation to writing proficiency. Joy Reid's 1992 study was one of many earlier attempts to analyze cohesive

devices in NNS writing by a computation tool. She analyzed 768 essays written by students from four separate language backgrounds: Arabic, Chinese, Spanish, and English.

The essays comprised two topic types and two further topics per type. In the first topic type, students could write an argumentative essay on the prompts of either space or leisure. In the second topic type, they could write a descriptive essay about a given chart, with choices on prompts of either farming or the continent. In the study, Reid (1992) examined four language variables associated with the concept of cohesion: pronouns, simple coordinate conjunctions (e.g., and, but, yet), subordinate conjunctions (e.g., when, while, which, before), and prepositions.

The analysis revealed that the four language groups significantly differed in the use of all four cohesive devices. Specifically, English speakers used far fewer pronouns and coordinate conjunctions than the other three language writers. However, in the variable of prepositions, English speakers used prepositions more frequently than other language groups. The analysis of subordinate conjunctions reveals puzzling results. Chinese students tended to use a higher percent of subordinate conjunction openers than English speakers.

Reid argued that such differences in language variables are attributed to NNS's first language interference and lack of rhetorical knowledge of English writings. For example, NNS's might not be aware of NS writers' rhetorical strategy to use fewer pronouns in order to express formality and distance to self; however, they tend to use more prepositions in formal writing in order to expand the prose's size and complexity.

Guobing Liu (2013) examined more specific use of cohesive devices by Chinese EFL learners, focusing on the use of linking adverbials in Chinese students' speaking and writing. His learner corpora include Chinese Learners' English Corpus and the College Learners' Spoken English Corpus, while he used Louvain Corpus of Native English Essays

and London-Lund Corpus as the NS control corpora.

In the analysis of 103 linking adverbials as a whole, Liu found that Chinese EFL learners tend to use considerably more linking adverbials than their NS counterparts both in speaking and writing. In-depth analysis of the top 15 linking adverbials in writing showed different linguistic patterns. Chinese EFL learners overused four adverbials (*so, in fact, then, of course*) but underused 10 adverbials (*anyway, also, well, indeed, though, actually, still, yet, even, therefore*) than NS speakers. One adverbial (*finally*) showed a similar pattern. The difference was greater in the overused adverbials.

Chinese EFL learners' overusing tendency is stronger in speaking than in their writing. They overused 13 linking adverbials (*so, also, well, of course, therefore, then, even, yet, still, finally, in fact, indeed, and anyway*) while only two (*though*, *actually*) were under-used. In particular, Chinese EFL learners' dependency on two adverbials (*so* and *also*) was salient, taking more than one half of the total frequency.

In comparison with speaking and writing, Chinese EFL learners and NS's demonstrated a contrasting pattern. Chinese EFL learners used the top 15 adverbials more frequently in speaking than in writing. In contrast, NS's used fewer linking adverbials in speaking than in writing. The difference might be attributed to the fact that more than half of the top 15 adverbials used by Chinese EFL learners are spoken register sensitive while most of NS's top15 adverbials are written register sensitive.

As for the difference in the use of linking adverbials between Chinese EFL learners and NS's, Liu suggested five possible explanations. Firstly, Chinese EFL learners used such linking adverbials as *first*, *for example*, *so*, *then*, *of course* and *in fact* because of mother tongue transfer. According to Liu, there are equivalent Chinese words for frequently used adverbials. Secondly, Chinese EFL learners might use enumerative adverbials such as

first(ly), *second*(ly), *third*(ly) and *finally* and many appositive linking adverbials such as *for example* because current Chinese/English writing instruction encourages such a writing pattern. Thirdly, Chinese EFL learners tend to use adverbials commonly used in writing also in speaking, because they might lack of register awareness. Fourthly, Chinese EFL learners overly use certain adverbials because they do not fully recognize the subtle differences in the meaning of other adverbials, hence they use more familiar adverbials repetitively. Lastly, Liu argued that the lower use of certain corroborated linking adverbials might be the writers' attempt to express depersonalization in writing.

The final study in this section was conducted by Crossley, Kyle, and McNamara (2016) who explored NNS's use of cohesive devices in relation to writing quality and traced their process of development over time. They collected each NNS university-level students' essays at the beginning, middle, and the end of their English course. Each student chose from one of two topics. The final corpus was 171 essays from 57 writers. The essays were rated by expert raters on five analytical features: content, organization, vocabulary, language use, and mechanics.

On the comparison of overall writing quality, Crossley et. al reported that the students' writing quality improved over time. In particular, their scores from midpoint were significantly higher than in initial essays; however, their scores were not significantly different from those in the final essays.

In addition, the study reported that students showed a growth in the use of cohesive devices over time. The increased cohesive devices are noun overlap between paragraphs, repeated content words and function words between sentences, paragraph and text, positive connectives and noun synonyms.

Crossley et al. further examined whether cohesive devices could predict writing

quality. They find that four cohesive indices are strong indicators of writing proficiency, explaining a 36 % variance of the human rating. The devices are adjacent overlap two paragraphs (function words), adjacent overlap two sentences (function words), adjacent overlap two paragraphs (pronouns), and pronoun-to-noun ratio. However, repetitive function words, incidence of coordinating conjunctions, and sentence overlap of pronouns were all negative predictors of writing quality.

The findings of this study suggest that NNS writing quality is closely related to the occurrence of cohesive features. It shows gains in certain cohesive devices as writing improves. Moreover, the study supports the notion that the use of certain cohesive devices can predict writing quality with a great deal of accuracy.

NNS use of discourse connectors

The concerns regarding NNS students' use of the cohesive devices have led to the study of many comparative studies on their use of sub-level category, especially the discourse connectors. For example, Crew (1990) investigated the misuse and overuse of connectors by analyzing the writing from ESL students at The University of Hong Kong. His research discovers that the connector *on the contrary* is often misused and argues that language composition textbooks may contribute to such misuse.

In a subsequent study on the use of connectors, Bolton, Nelson, & Hung (2002) compared the writing of university students in Hong Kong and Great Britain, using the Hong Kong subset data from the International Corpus of English in Hong Kong (ICE-HK) and the British data from the International Corpus of English (ICE-GB) respectively. They also used the list of the connectors derived from the academic writing in the ICE-GB subset as the norm of the study and compared the use of connectors with those of college students in Hong Kong and Britain. Their findings reveal that both Hong Kong and British students tend to

overuse connectors when compared to academic writers, but they show a significant difference in their choice of the connectors. Hong Kong students overly employ *so*, *and*, *also*, *thus*, *but*, in the order of occurrence while British students frequently use *however*, *so*, *therefore*, *thus*, *furthermore*.

In contrast to Hong Kong and British students, the academic writers use connectors less in their writings (Bolton, Nelson, & Hung, 2002). For example, even the top frequently used connector among the academic writers, *however*, is used 20.4 times per 1,000 sentences while it is used 23.6 times and 40.9 times by Hong Kong and British students respectively. Their research demonstrates that Hong Kong students show similar language patterns to that of the novice writers who are native speakers by frequently employing similar connectors; however, they choose different connectors compared to the British students. While the authors' research provides an important insight into the learners' language patterns, they used the subsets of two separate corpora for the writings of Hong Kong, British, and academic writers. Therefore, it is difficult to generalize the findings since the topics of each corpus varied.

In summary, there is substantial evidence that Asian students use sentence connectors in different ways from native English speakers. That is, they employ certain connectors more or less often than native speakers. The question is whether such contrasting use of connectors will show differing results, depending on their English knowledge and the topic. Another question is whether the results are similar among students in separate Asian countries.

Chapter 3: Methodology and computational tools

Material

The current study utilizes the written essay module of the ICNALE (the International Corpus Network of Asian Learners of English). It is one of the largest learner corpora available to this date, totaling 1.3 million words (see Table 1). The corpus was originally developed by Dr. Shin Ishikawa, Kobe University in Japan. Since its inception in 2011, Dr. Ishikawa has continued to enlarge its size by gradually adding more countries, and now it contains more than 10,000 topic-controlled speeches and essays produced by college students in ten Asian countries as well as English native speakers. Currently the corpus includes four modules: Spoken Monologue, Spoken Dialogue, Written Essays, and Edited Essays. Among the available modules, the present study only utilizes the written essay module.

Participants in the ICNALE (the International Corpus Network of Asian Learners of English)

Dr. Shin Ishikawa, Kobe University in Japan, gathered data of a total of 2,600 college students including graduate students from countries in both ESL (Hong Kong, Pakistan, the Philippines, and Singapore) and EFL regions (China, Indonesia, Japan, Korea, Taiwan, and Thailand). Additionally, native speakers of English participated in this project. Two hundred native speakers' data were collected, too. One hundred speakers were college students while the others were working professions such as instructors, translators, writers, and professors. Their nationalities were U.S. (114), Britain (28), Canada (28), Australia (17), and New Zealand (13).

Controlled writing conditions

The essays of ICNALE were strictly controlled for the factors that might influence language in order to make homogeneous data across corpus. More specifically, the

participants were given identical instructions. They were to write two essays stating their opinions with supporting details about two topics: (A) *It is important for college students to have a part-time job,* hereafter "part-time prompt" *and* (B) *Smoking should be completely banned at all the restaurants in the country,* hereafter "smoking prompt". The participants also were required to use MS Word or a similar word processor and run a spell-check before completion. No dictionary or other reference tools were allowed. The essay should be from 200 to 300 words. The participants were given 20 to 40 minutes per essay.

The following table 1 summarizes the corpora used for the present study.

[Table 1]

Key corpora facts

Country	Total participants	Total words in Part-time prompt	Total words in Smoking prompt
Chinese	400	96577	92766
Hong Kong	100	23848	23054
Indonesia	200	47100	46062
Japanese	400	89320	87817
Korean	300	68480	66485
Pakistan	200	47247	47430
Philippine	200	50469	48500
Singapore	200	49688	48186
Thailand	400	90381	89676
Taiwan	200	46772	44615
English speakers	200	44825	45051
Total	2800	654707	639642

In addition to writing conditions, ICNALE is also controlled for writers' English proficiency level. Prior to the writing task, the participants were required to take the English vocabulary size test (VST) (Nation & Beglar, 2007). Those who have taken a standard test such as TOEIC and TOEFL reported their scores. Besides, they completed the questionnaires surveying their exposure to English. Combined with all these factors, the participants were

divided into four proficiency level (B2, B1_1, B1_2 and A2 as B2 is the highest and A2 is the lowest). In this study, I will call A2 as the beginner, B1_1 as the intermediate low and B1_2 as the intermediate high, and B2 as the advanced for convenience.

Instruments of the Study

      AntConc. The current study uses *AntConc* (Version 3.5.7), a free corpus analysis tool. It was developed by Professor Laurence Anthony at Waseda University, Japan, originally for the use of technical writing instruction (Anthony, 2006) and can be downloaded at his website: *http://www.antlab.sci.waseda.ac.jp/antconc_index.html*. Since its first release in 2002, the software has been updated 19 times and the latest version is *AntConc 3.5.7*. Users don't need to install this on their computers and can launch the program by simply double-clicking an executable file even from a USB memory stick. Additionally, the current version supports other languages such as Japanese and Korean and the different operating systems such as Windows, Macintosh, and Linux.

      For its user-friendly environment, *AntConc* has been widely applied in diverse fields of second language research. For example, Flowerdew (2015) demonstrated its use in writing instruction by conducting multi-step workshops to postgraduate science and engineering students and offering helpful tips in composing their theses. In another study, Yunxia, Min & Zhou (2009) reported the beneficial use of *AntConc* in English vocabulary teaching and learning environment while Chang and Kuo (2011) demonstrated its usefulness in more genre-specific academic instruction in the academic field of computer science.

      The most notable use of *AntConc* is in the field of contrastive corpus research. Römer and Wulff (2010) used the software in analyzing the occurrence of *this* depending on the students' academic disciplines and their school years in the large corpus of the Michigan Corpus of Upper-level Student Papers (MICUSP). More recently, it is used in the study of

Martinez (2018) in exploring frequency and range of word lists used by international scientists as compared to the native speakers of English in their published journals.

AntConc includes several linguistic analysis tools such as the Concordancer, Search Term Distribution Plot, Original File View, Word Clusters/Lexical Bundles, Word list and Keyword lists. Of these tools, this study primarily uses the Concordancer.

The Concordancer tool is the central feature in *AntConc* as it is displayed into the main screen. This tool searches and retrieves a specific word or a phrase from a given text (or sets of text) and displays the search results. The results are highlighted in the main screen along with some context left and right of them. This presenting format of providing information is called "KWIC": Key word in Context. Additionally, the Concordancer tool in *AntConc* provides the raw frequency of search results. Therefore, it allows the users to find out how frequently words or phrases are used in the target corpus.

The use of Concordancer in *AntConc* is very straightforward. When users download from the website and open *AntConc*, they will see that the concordance tab is already displayed by default. So, they need to select the file tab and open the file or directory of their choice. The chosen file(s) will be displayed in the left window of corpus files. Then, they need to enter a search word (or phrase) in the box under the main window and click the 'Start' button. By default, *AntConc* concordance searchers are case-insensitive so they need to click on 'Case' box next to 'Search Term' if they want a case-sensitive search. Figure 1 shows the result of the connector *finally* in Chinese students' essay on Part-time job using *AntConc*.

Due to the limited size of the computer screen, only a certain number of contextual words is displayed in each search result. Depending on the type of analysis, it may be necessary to see more contextual views. If users would like to look at whole text of the search result, for example *finally* in Figure 1, they need to click on the blue highlighted *finally* in the

concordance line. Then, the 'File View' tab opens and displays a whole text with 'finally'

highlighted (see Figure 2). They can go back to concordance result window by clicking on

concordance tab.
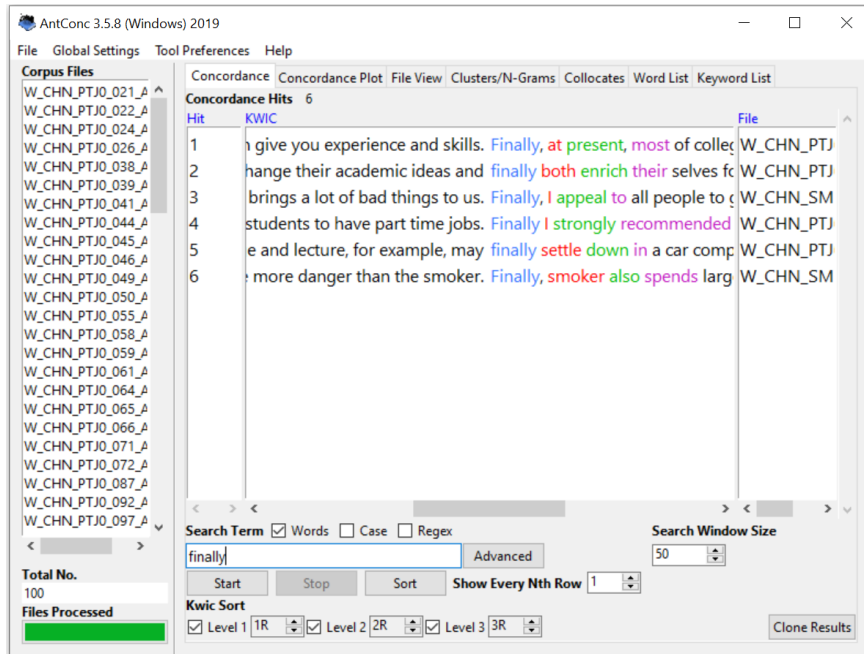
[Figure 1] *AntConc* concordance of the word "finally" in the Chinese subcorpus
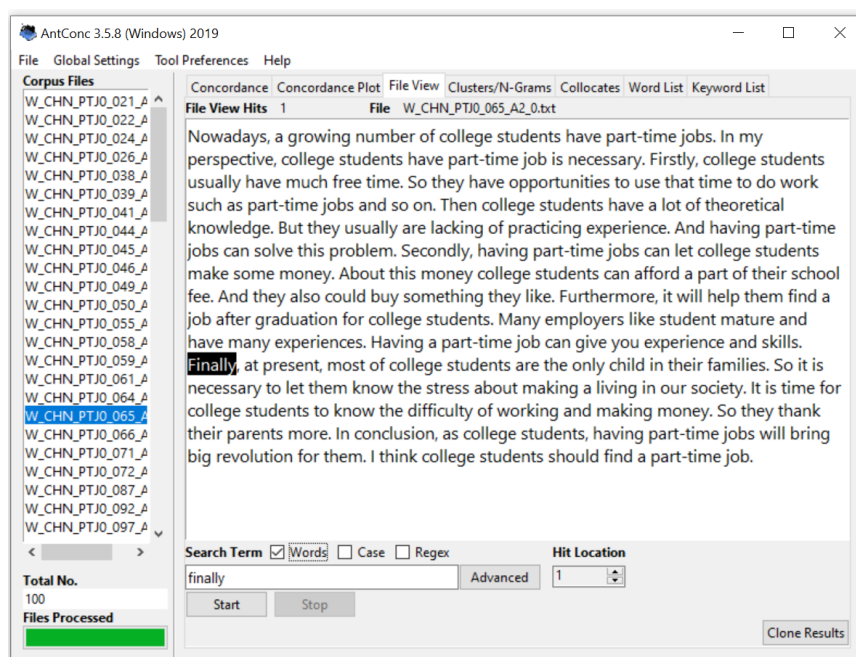
[Figure 2] *AntConc* concordance of the word finally in the Chinese subcorpus

Coh-Metrix. Coh-Metrix is an automated tool developed at the Institute for Intelligent Systems at the University of Memphis. This textual tool provides linguistic indices that measure important textual characteristics at multiple levels such as text cohesion, text sophistication, and text readability (Crossley et al., 2011) as well as the target text's descriptive features such as number of words and sentences. Coh-Metrix includes over 600 indices of linguistics textual features (Crossley & McNamara, 2010).

A number of studies in second language research has validated the power of Coh-Metrix in analyzing textual characteristics, most notably textual cohesion and linguistic sophistication indices (Crossley & McNamara, 2010) and L2 lexical development indices (Crossley, Salsbury, & McNamara, 2009). Moreover, Cho-Metrix is used to differentiate linguistic features of diverse texts such as writings by Japanese, American and British Scientists (McCarthy et al., 2007), journal articles between Chinese and American scientists (Ye, 2013), and Ph.D. dissertations written by Iranian University students and English native speakers (Azadnia, Lotfi, & Biria, 2019). For this study, I chose the indices in Coh-Metrix that report descriptive characteristics of corpora.

Coh-Metrix software is available at the Institute for Intelligent Systems at the University of Memphis website, http://cohmetrix.memphis.edu/cohmetrixpr/index.html. Currently, version 3.9 is available on the site; however, the online version requires the user to enter each text on the window, not appropriate to process large data. Therefore, I contacted Dr. Zhiqiang Cai at the Institute for Intelligent Systems and received a downloadable Coh-Metrix software. So I was able to process a large file at a time.

In order to analyze the linguistic features of the corpora, I selected the following indices.

Number of words. This feature calculates the total number of words in a text.

Number of sentences. This feature calculates the total number of sentences in a text.

Connectives. This feature calculates the incidences of all connectives per 1000 words. The indices are offered in seven general classes: causal (*because, so*), adversative/contrastive (*although*, *whereas*), logical (*and, or*), temporal (*first, until*), additive (*and, moreover*). Additionally, there is a distinction between positive (*also, moreover*), and negative connectives (*however, but*).

Procedure

The present study aims to understand the linguistic features in college students from ten Asian countries, as compared to English native speakers, especially on the use of sentence connectors. To better understand the textual features of corpus, I take two steps using two different tools to highlight the linguistic differences. In the first step, I use Coh-Metrix and get a descriptive feature of each sub-corpus such as the total number of sentences in a text, its mean in a corpus, the total number words in a sentence and its mean in a text. In addition to general descriptive feature, I use Coh-Metrix to identify overall connectors use and seven categories of connectors (causal, contrastive, additive, logic, temporal, positive, negative).

Followed by the general descriptive data of corpora produced by Coh-Metrix, I used the concordance module of *AntConc* to further investigate the use of sentence connectors. In doing so, I followed the protocol of Bolton et al. (2002). Firstly, I made the list of 118 connectors adopted from Quirk et al. (1993). Table 2 shows the full list of connectors by the categories used in this study. Secondly, I examined the type and frequency of the connectors in the reference corpus (i.e. English speaker's corpus). Lastly, I then examined the learner corpus and compared the results with the data from the reference corpus.

[Table 2]

List of connectors by categories adopted from Quirk et al. (1993)

Categories	Connectors
Enumeration (addition, equative, reinforcing)	additionally, alternatively, also, and, besides, by the same token, further, furthermore, in addition, in the same way, likewise, moreover, neither, nor, on top of that, or, or else, similarly, too, as a final point, at this point, finally, first/firstly, first of all, for a start, for another thing, for one thing, from now on, henceforward, hitherto, in the first place, in the second place, last/lastly, last of all, next, second/secondly, then, third/thirdly, to begin with, up to now, for example, for instance, in a word, in other words, namely, more precisely, that is, that is to say, to put it another way, what is to say, actually, as a matter of fact, as it happens, at any rate, at least, in actual fact, in any case, indeed, in either case, in fact, in reality, to tell the truth
Summation	all in all, anyway, briefly, in conclusion, in short, in sum, in summary, overall, to conclude with, to get back to the point, to resume, to summarize, to sum up
Result/Interference/Cause	accordingly, arising out of, as a consequence, as a result, aside from this, because, consequently, for this purpose, for this reason, hence, in consequence, in that case, in this respect, in such an event, on account of, on this basis, or, otherwise, so, then, therefore, thus, under the circumstances, with this in mind, with this intention
Contrast/Concession	anyhow, but, by comparison, by contrast, by way of contrast, conversely, despite this, however, in contrast, in spite of, instead, nevertheless, nonetheless, on the contrary, on the other hand, rather, still, though, yet
Transition	by the way

In a process to extract the raw frequency of connectors in the corpus, the present study faced a methodological problem. In English, many words in the connector list have various grammatical roles other than connecting a sentence or phrase. For example, the word *so* in English can be used in different grammatical roles. In a sentence like "*It is common for a young man to bear a loan from the bank for some reason and support his family or himself. So a part-time job even may entertain the person as well as bring in money, if the job is just his or her interest* (Chinese_part-time_015_B1_2), *so* is used as a sentence connector expression consequence. However, in a sentence like "*If you really have got a lot of works to deal with, it's better to spend all your efforts in your own studying, not even thinking about searching a part-time job, for that you can learn most things in the shortest time in school, at least I think so* (Chinese_part-time_209_B1_1)," *so* is used as a pronoun that refers back to the writer's argument.

In order to eliminate any other use of the connectors, the researchers have to manually scrutinize each incidence and determine its use. Such a process can be attainable if the corpus size is small. However, the present study utilizes a large number of texts. The incidence of *so* alone totals to 1,306 in Chinese students' corpus. Consequently, it is difficult to examine each incidence in the current corpus. To address the problem, the present study considers only when the connectors begin with a capital letter, collecting the incidences when they are at the beginning of sentences.

Once *AntConc* extracted all the incidence of certain connectors at the beginning of sentences, I looked through each incidence to ensure that it connects sentences and eliminate it if it is not at the beginning of a text. The total frequency was then normalized by dividing it with the total number of sentences in the corpus. Then, the learner corpora were categorized by the students' countries and by their English levels and analyzed by the occurrence of the

connectors through *AntConc*. The findings were normalized by the total number of sentences in the corpus. Finally, I compared the two results from the reference and learner corpora and presented the commonalities and differences in terms of the type and frequency of the discourse connector. In all instances, the frequencies per sentences are multiplied by 1,000 to eliminate very low figures. More in-depth discussion was conducted for the most frequently used connectors across the corpora.

Chapter 4: Results

The current study explores the linguistic similarity and difference of sentence connector use between Asian students from 10 different countries and English speakers. Specifically, it asks three research questions: 1) Is there a similarity in the use of connectors between the writing of college students in 10 Asian countries and native speakers (NS) of English, in terms of frequency and choice of connector? If there is, how do they employ the connectors similarly? 2) Is there a difference in the use of connectors between the writing of college students in 10 Asian countries and native speakers of English in terms of frequency and choice of connectors? If so, how do they employ the connectors differently? 3) Is there a difference in the use of connectors between the writings of college students within 10 Asian countries depending their English proficiency level and their nationalities? If so, how do they employ the connectors similarly and differently across English level and nationality?

To address the questions above, I present the results in three sections. The first section will provide a basic textual characteristic of corpora using the Coh-Metrix computational tool. In particular, it will provide data on the mean number of words, sentences and words per sentence. This analysis serves to help interpret the sentence connector results outlined in the second section of the paper.

The second section will report the results of sentence connector use in the Asian students' and English speakers' writings again using the computational tool, Coh-Metrix. This analysis will identify the total number of all connectors between the writings of the English speakers and the Asian students. Additional analysis will be conducted on the occurrence of the connectors in seven sub-categories (causal, logical, adversary/contrastive, temporal, additive, positive, negative) in order to highlight how each group of writers uses the connectors similarly and differently. Additionally, I will conduct a comparative analysis on

connector use according to the Asian students' origin country and their four English proficiency levels (beginner, intermediate-low, intermediate-high and advanced).

In analyzing data generated by Coh-Metrix, I use SPSS *ver*. 26 to conduct a pairwise comparison using an independent t-test when comparing English speakers and the Asian students in order to determine whether the difference in the descriptive data is statistically significant. In an analysis with more than two groups (e.g. students' nationalities and proficiency levels), I conducted MANOVA (Multivariate Analysis of Variance) and followed up the significance with Tukey's HSD post-hoc tests. The results are presented by two significant levels. If the p-level is lower than 0.05, it indicates that the result has less than a 5 % probability that the difference is not significant. That is, there is strong evidence against the difference being just random; therefore, the difference is significant by statistical calculation. If the p-level is lower than 0.001, it indicates that the result has less than one in a thousand chance of the difference being just random. In other words, if the p-level between two groups is less than 0.001, the difference between two groups is extremely significant by statistical calculation. When multiple comparison is conducted, I adjusted the alpha level by dividing it with the number of comparisons in order to protect again Type-1 error.

The final section of this chapter will use a concordance module of AntConc to offer more in-depth analysis as to which specific connectors are often used in a respective corpus. It will describe the most frequently used connectors of choice in the English speakers and Asian students. The results will be compared with that of the Asians students' origin country and English proficiency level. In this section, their preferred choice of connectors and frequency will be discussed. By combining the results from two computation tools, I aim to provide better understandings on how differently and similarly the Asian students use the sentence connectors as compared to their English counterparts.

Textual characteristic of corpora

Using Coh-Metrix, this section will present the textual characteristic of the Asian students' writing, such as the mean number of words, sentences, and words per sentence, to understand the corpus composition features as compared to that of the English speakers. The current corpora comprise writings of only one paragraph; therefore, the attribute of the number of words in a text is equivalent as the number of words per paragraph in this study and the number of sentences in a text as the number of sentences per paragraph correspondingly.

[Table 3]

Descriptive analysis on the writings of English speakers and the Asian students

	English	Asian	P value
Number of words	224.69(24.05)	231.63(30.99)	<.001
Number of sentences	9.29(2.81)	14.62(4.36)	<.001
Number of words per sentence	26.10(5.99)	18.11(11.84)	<.001

Note: Decimals are rounded off to two decimal places.
Standard deviations are in parenthesis.

Coh-Metrix analysis between the English speakers and the Asian students reveals that the two groups have significant differences in textual characteristics (see Table 3). On the average of 224.69 words per text, the English speakers use significantly fewer words than the Asian students (*M*=231.63, *SD*=30.99), t(5598)= -4.38, p<.001, *d*=0.25. Moreover, the Asian writers use a fewer number of words per sentence (*M*=18.11, *SD*=11.84), therefore, more sentences per text than the English speakers, t(5598)=-24.13, p<.001, *d*=1.46. Moreover, in a timed writing context, the Asian students using more words and more sentences in a text produce sentences that are significantly shorter than those in the English speakers' writing, t(5598)=13.36, p<.001, *d*=0.85.

Coh-Metrix analysis on the frequency of the connectors

In this stage of the analysis, Coh-Metrix is used to investigate the similarity and difference of the connector use in terms of frequency and function between the English speakers and the Asian students. More specifically, this section explores whether the Asian students use the connectors differently from their English counterparts and to what extent the difference lies across their origin countries and English proficiency level.

Using Coh-Metrix, this section provides the quantitative comparison between the Asian students and the English speakers on the total occurrence of connectors per 1,000 words. In this analysis, Coh-Metrix includes all the connectors regardless of their position in the sentence (although AntConc analysis primarily targets connectors at sentence-initial position). Furthermore, I will explore the occurrence of the connectors in the seven sub-categories (causal, logic, contrastive, temporal, additive, positive, negative) to investigate how the Asian students use the connectors differently or similarly when compared to the English speakers. With the generated data, I conduct MANOVA (Multivariate Analysis of Variance) with Tukey HSD post-hoc test in order to determine whether the difference is indeed significant.

English speakers vs the Asian students.   Table 4 shows the results on the mean frequency of all connectors and seven sub-categories per 1,000 words between the English speakers and Asian students. The results indicate that the English speakers use additive connectors (e.g. *and*, *moreover*) most frequently followed by positive connectors (e.g. *also*, *moreover*). On the other hand, the Asian students rank the positive connectors at the top most frequently used connector type, followed by logical connectors (e.g. *and*, *or*).

[Table 4]

Connector use of English speakers and the Asian students

Connectors	English	Asian	P value
All connectors	104.82(20.13)	102.55(21.45)	.041
Causal	44.93(17.18)	41.03(15.77)	.000*
Logic	38.47(23.12)	60.12(17.59)	.000*
Contrastive	18.35(9.43)	19.02(10.76)	.225
Temporal	17.19(10.89)	15.55(9.91)	.002*
Additive	69.53(26.17)	46.67(14.66)	.000*
Positive	55.28(43.44)	91.08(21.22)	.000*
Negative	17.67(9.87)	13.59(8.53)	.000*

Note: Decimals are rounded off to two decimal places.
Standard deviations are in parenthesis.* significant at .006

To prevent Type 1 error in conducting multiple t-test, I made a Bonferroni correction and used an alpha level of .006 for the sub-category analysis. The analysis revealed that the Asian students (*M*=102.55, *SD*=21.45) didn't use differently all connectors combined as compared to their English counterpart (*M*=104.82, *SD*=20.13), t(5598)=2.05, *p*=0.41, *d*=0.11 In the analysis of sub-categories, the Asian students use significantly more connectors in logical and positive functions, which is an indicator of lower level writing (Crossley & McNamara, 2011). However, they used significantly fewer connectors in causal, temporal, additive and negative functions than the English speakers. Furthermore, the analysis found there is no significant difference in the use of contrastive connectors between the Asian students and the English speakers, t(5598)=-1.21, *p*=.225, *d*=0.067. Accordingly, the results suggest that the Asian students could compose more native-like writings if they use a higher number of causal, temporal, additive and negative connectors and rely less on logical and positive connectors in creating a link between two propositions.

English speakers vs the Asian students depending on their origin country. For the occurrence of connectors per 1,000 words, there was a statistically significant difference in

the use of connector types based on the Asian students' respective country, $F(80, 35412) = 60.32$, p <.001; Wilk's $\Lambda = 0.445$, partial $\eta^2 = .096$. The relationship with the type of connectors and the nationalities further was analyzed using MANOVA. The multivariate result was significant for the variables of all connectors, causal, logic, contrastive, temporal, additive, positive, and negative connectors, indicating that depending on the countries the Asians students show a significant difference in the use of connector type compared to the English speakers (see Table 5) at p <.006 (with a Bonferroni correction).

[Table 5]

*Differences in the type of connectors based on nationalities*

Variables	F	Sig	Eta squared
All connectors	31.46	0.00	0.05
Causal	53.77	0.00	0.09
Logic	83.07	0.00	0.13
Contrastive	9.32	0.00	0.02
Temporal	22.90	0.00	0.04
Additive	93.89	0.00	0.14
Positive	114.05	0.00	0.17
Negative	15.75	0.00	0.03

Followed by MONOVA test, I conducted a Tukey's post-hoc tests to determine the pairwise difference between the English speakers and the Asian students' respective countries in the use of connector types. To protect against Type 1 error, I adjusted the alpha level at .005 (with a Bonferroni correction).

Taking into consideration all types of connectors, only Chinese and Hong Kong students use significantly fewer connectors as compared to the English speakers ($M=104.82$, $SD=20.13$). Especially, Hong Kong students ($M=94.60$, $SD=18.54$, $p<.001$) use the fewest connectors in their writings. On the other hand, a majority of Indonesia, Japanese, Korean, Pakistan, Philippian, Taiwan, and Thailand students do not show a significant difference in

the overall use of the connectors.

In pairwise comparison of the connector categories (see Table 6 and Figure 3), the Asian students show a homogeneous pattern in four functions. In the logical and positive connectors, writers from all 10 Asian countries use significantly more connectors than the English speakers. Indonesian students use the greatest number of logical connectors (M=66.77, SD=20.20, p<.001) and Thailand students use the most positive connectors (M=98.53, SD=22.70, p<.001). In the additive and negative connectors, the Asian students use fewer connectors across their origin countries than the native English speakers. Especially, Hong Kong students use the fewest number of additive (M=40.25, SD=13.06, p<.000) and negative connectors (M=11.91, SD=6.65, p<.000), which shows the greatest discrepancy with the English speakers.
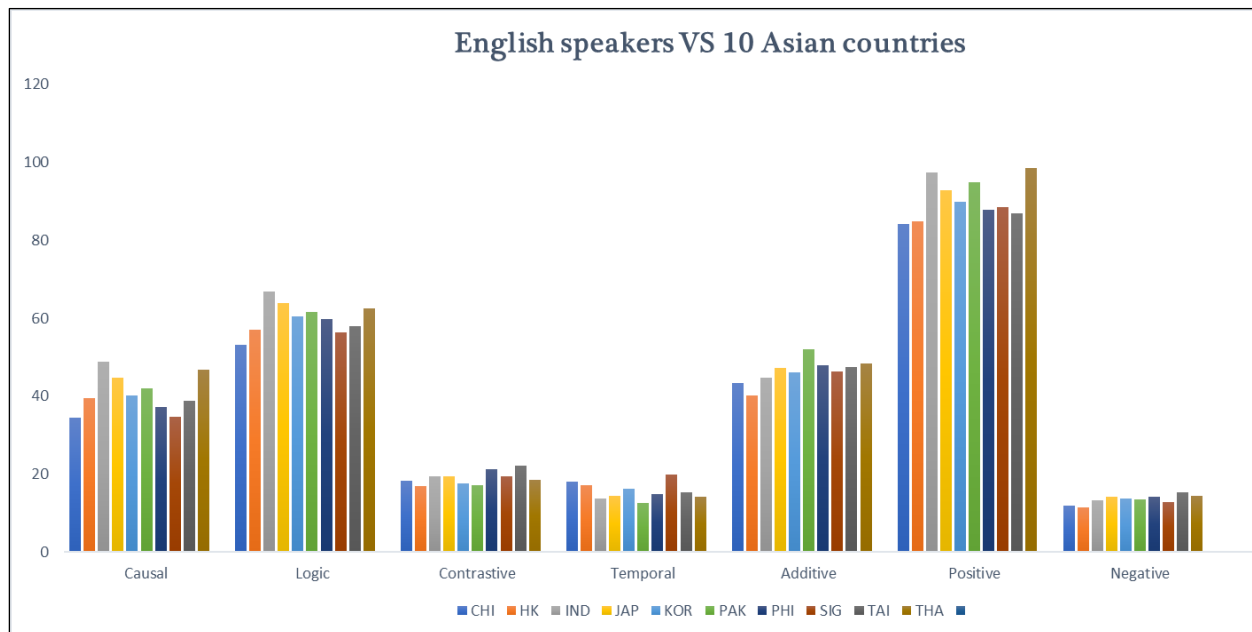
[Figure 3] The occurrence of the connectors per 1000 words by Asian students' origin country

[Table 6]

The occurrence of the connectors per 1000 words by Asian students' origin country

	ENG	CHI	HK	IND	JAP	KOR	PAK	PHI	SIG	TAI	THA
all connectors	104.82 (20.13)	*95.19 (18.38)	*94.60 (18.54)	109.58 (22.44)	104.55 (19.6)	101.01 (21.09)	106.42 (25.21)	99.81 (20.41)	100.02 (18.41)	99.92 (20.52)	109.58 (22.93)
Causal	44.93 (17.18)	*34.56 (12.73)	*39.44 (15.00)	48.75 (17.83)	44.72 (14.37)	*40.15 (15.04)	42.00 (16.05)	*37.19 (14.94)	*34.84 (13.09)	*38.74 (13.82)	46.71 (17.22)
Logic	38.47 (23.12)	*53.11 (14.29)	*57.08 (16.54)	*66.77 (20.20)	*64.00 (15.52)	*60.57 (17.22)	*61.69 (20.48)	*59.84 (17.83)	*56.49 (15.46)	*58.00 (15.92)	*62.61 (18.82)
Contrastive	18.35 (9.43)	18.24 (9.05)	16.92 (9.19)	19.47 (11.77)	19.55 (10.78)	17.73 (10.20)	17.22 (11.41)	*21.33 (11.66)	19.51 (10.13)	*22.13 (10.67)	18.48 (11.51)
Temporal	17.19 (10.89)	18.03 (9.46)	17.13 (10.66)	*13.65 (10.14)	*14.36 (9.01)	16.18 (10.37)	*12.64 (9.80)	14.83 (9.42)	19.88 (9.70)	15.33 (10.02)	*14.13 (9.60)
Additive	69.53 (26.17)	*43.47 (13.05)	*40.25 (13.06)	*44.69 (14.4)	*47.36 (14.08)	*46.2 (14.23)	*52.06 (17.51)	*48.00 (14.72)	*46.43 (13.27)	*47.39 (13.69)	*48.50 (15.69)
Positive	55.28 (43.44)	*84.22 (18.60)	*84.91 (18.57)	*97.48 (21.62)	*92.83 (19.12)	*89.86 (20.50)	*94.83 (24.46)	*87.92 (20.84)	*88.55 (18.51)	*86.81 (20.51)	*98.53 (22.70)
Negative	17.67 (9.87)	*11.91 (6.65)	*11.55 (7.09)	*13.31 (8.91)	*14.18 (8.51)	*13.84 (8.73)	*13.45 (9.68)	*14.20 (8.59)	*12.84 (8.09)	**15.27 (8.01)	*14.35 (9.62)

Note: Decimals are rounded off to two decimal places.
*Standard deviations are in parenthesis. * indicates significant difference with English speakers at .005*

In the attribute of causal, contrastive and temporal connectors (see Table 6 and Figure 3), the Asian students show a heterogeneous pattern, as compared to the English speakers. In the attribute of causal connectors (e.g. *because*), Chinese, Hong Kong, Korean, Philippine, Singapore and Taiwan students use significantly fewer than the English speakers at .005. On the other hand, the results reveal that Indonesia, Japanese, Pakistan, and Thailand students show no difference in the use of the causal connectors compared to the English speakers. In a similar way, Philippine and Taiwan students use significantly more contrastive connectors compared to the English speakers. However, the rest of the Asian countries does not differ in the use of contrastive connectors in comparison with the English speakers.

In temporal connectors (e.g. *when*, *before*), Indonesia, Japanese, Pakistan, and Thailand students use significantly fewer connectors compared to the English speakers (M=17.19, SD=10.89, p <.005). Chinese, Hong Kong, Korean, Philippian, Singapore and Taiwan students do not show a difference in the use of temporal connectors compared with the English speakers.

Taken together, the Asian students show a great variance on the connector use in the all connectors, causal, contrastive and temporal connectors, depending on their origin countries. However, they consistently show an overuse of the logical and positive connectors and the underuse of the additive and negative connectors across their respective countries compared to the English speakers.

English speakers vs the Asian students depending on their proficiency level.   A MANOVA test was conducted in analyzing the connector use between the English speakers and the Asians students depending on their English proficiency levels (see Table 7). To protect against Type-1 error, I adjusted the alpha level at .013. The multivariate results show that there is a statistically significant difference in the use of connectors based on the English proficiency

level, $F(32, 20609.12) = 110.97$, p<.001; Wilk's $\Lambda = 0.57$, partial $\eta2 = 0.14$. Univariate F tests

show that the English proficiency level has a significant effect on all sub-categories of

connectors except when all connectors are combined.

[Table 7]

Differences in the type of connectors based on English proficiency levels

Variables	F	Sig	Eta squared
All connectors	2.31	.055	.00
Causal	14.49	.000*	.01
Logic	135.49	.000*	.09
Contrastive	3.91	.004*	.00
Temporal	20.38	.000*	.01
Additive	196.93	.000*	.12
Positive	217.86	.000*	.14
Negative	21.07	.000*	.02

Note: Decimals are rounded off to two decimal places.
* indicates significant difference with English speakers at .013

Followed by MONOVA test, I conducted a Tukey's post-hoc tests to determine the

pairwise difference between the English speakers and the Asian students' English proficiency

level in the use of connector types (see Table 8). The alpha level was again adjusted at .013 (with

a Bonferroni correction).

In the overall use of connectors depending on the English proficiency level, the Asian

students tend to use fewer connectors as their English proficiency level increases (see Table 8

and Figure 4); however, the difference was not statistically significant as compared to the

English speakers.

Subsequent analysis on the use of temporal connectors reveals that the Asian students

gradually follow the norm of the English speakers as their English improves (see Table 8 and

Figure 4). For example, the students in the beginner level use significantly fewer (*M*=14.01,

SD=9.85, *p*<.001) than their English counterpart (*M*=17.19, *SD*=10.89). However, they tend to use more temporal connectors in the advanced level (*M*=18.52, *SD*=10.39) which show no difference with the English speakers (*p*=.286). Conversely, the Asian students in the advanced level successfully follow the norm of the English speakers in terms of frequency.
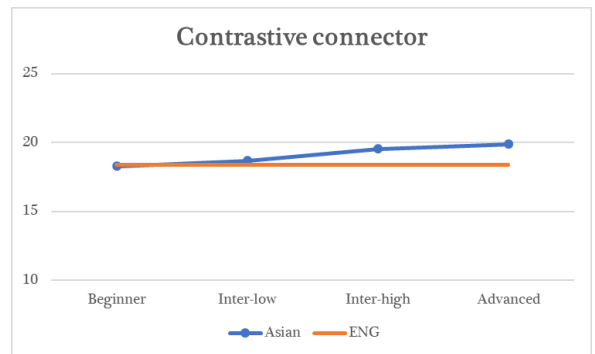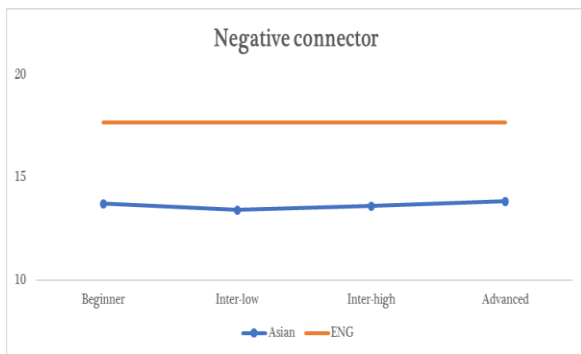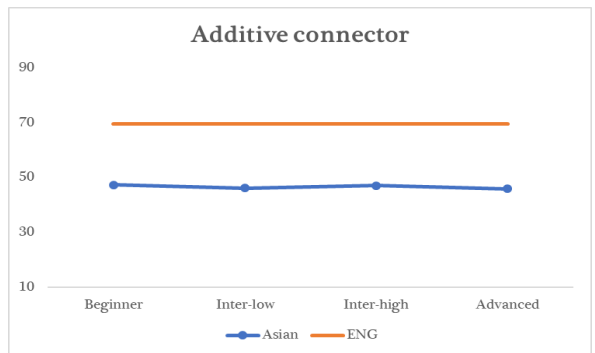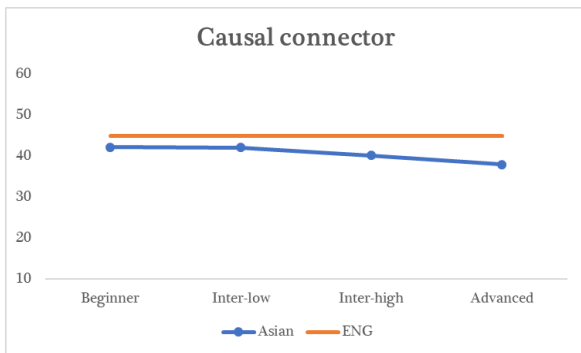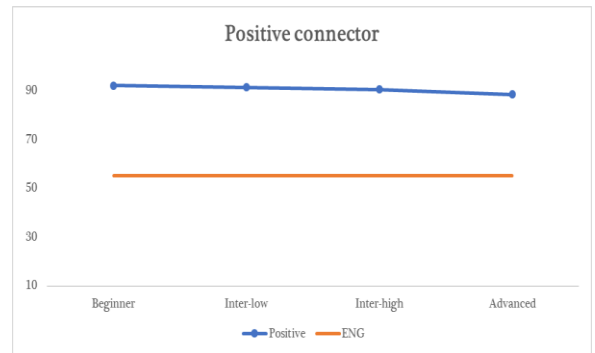
[Table 8]

The occurrence of the connectors per 1000 words by students' English proficiency level

	ENG	Beginner	Int-low	Int-high	Advanced
all connectors	104.82 (20.13)	103.36 (21.86)	102.75 (21.71)	102.39 (21.33)	100.72 (19.89)
Causal	44.93 (17.18)	42.17 (15.93)	**42.04 (16.14)	*40.18 (15.61)	*37.97 (13.93)
Logic	38.47 (23.12)	*60.90 (17.63)	*60.46 (17.60)	*59.75 (17.84)	*58.62 (16.30)
Contrastive	18.35 (9.43)	18.28 (11.26)	18.67 (10.66)	19.54 (10.71)	19.88 (10.16)
Temporal	17.19 (10.89)	*14.01 (9.85)	*15.16 (9.62)	16.01 (9.92)	18.52 (10.39)
Additive	69.53 (26.17)	*47.33 (15.05)	*46.18 (14.58)	*47.03 (14.71)	*45.85 (13.83)
Positive	55.28 (43.44)	*92.27 (21.34)	*91.54 (21.75)	*90.62 (21.14)	*88.56 (18.70)
Negative	17.67 (9.87)	*13.73 (9.34)	*13.42 (8.39)	*13.62 (8.31)	*13.85 (8.28)

Note: Decimals are rounded off to two decimal places.
*Standard deviations are in parenthesis. * indicates significant difference at .013*
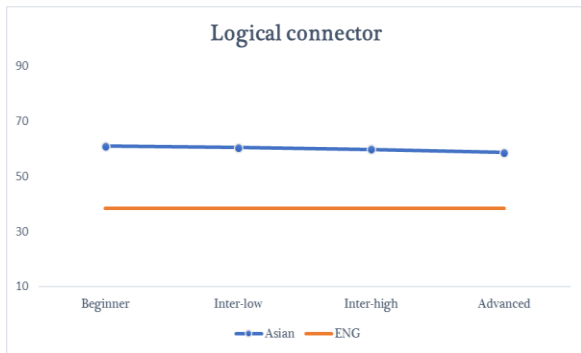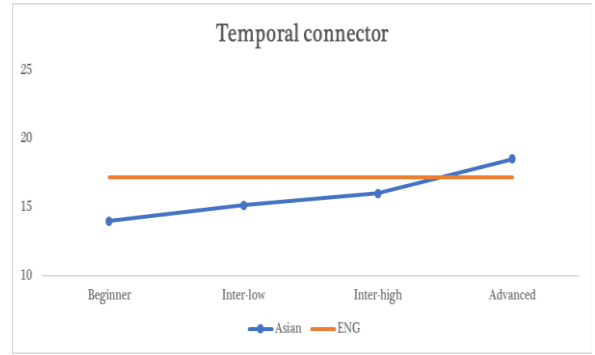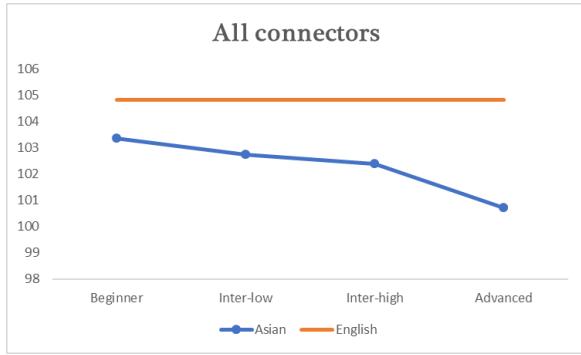
[Figure 4] The occurrence of the connectors by Asians students' English proficiency level vs English speakers

In the frequency of use of logical and positive connectors (see Table 8 and Figure 4), the Asian students in all levels show significantly more use than the English speakers. In the use of logical connectors, the beginner level Asian students use 60.90 incidences per 1,000 words, significantly higher than the English speakers (M=38.47). As their English improves, their mean of logical connectors slightly decreases but the difference is minimal within the level. Such tendency is also observed with positive connectors.

Similarly, the Asian students do not follow the model of the English native speakers on the attribute of additive and negative connectors as their English level increases (see Table 8 and Figure 4). They continue to use a significantly fewer number of connectors even though their English improves (p<.001). In the negative connectors, the beginner level students use only 13.73 incidences per 1,000 words on the average while the English speakers use 17.67. As their English progresses, the Asian students continue to use significantly fewer negative connectors p<.001). Lastly, the Asian students in all levels do not show significant difference in the use contrastive connectors compared to the English speakers.

Combined together, the connector analysis reveals that there is a difference in terms of frequency between the English speakers and the Asian students depending on their English proficiency levels. Their use of temporal connectors follows the model of the English speakers as their English progresses. However, they do not show much difference in the causal, logical, additive, positive and negative connectors in terms of frequency even as their English improves.

Coh-Metrix analysis provides valuable information as to the frequency difference of the connectors between the English speakers and the Asian students regardless of their position in the sentence. However, it doesn't highlight whether a respective group prefers to use specific connectors frequently or scarcely. Next we need to examine if there is a difference in terms of

connector choice in respective corpus.

AntConc analysis on the choice of connectors

While Coh-Metrix gives a bird's eye view of the connector use between the Asian students and the English speakers, it doesn't offer an in-depth analysis as to which specific connectors are frequently used in respective language groups or their English proficiency level. To address such a problem, I use the concordance module in AntConc, which extracts the incidences of specific phrases or words with immediate contexts in corpus.

In the first stage of the AntConc analysis, I process the selected 118 connectors (see Table 2 in Chapter 3) on the English speakers' and the Asian students' writings to extract their most used connectors respectively. The result includes the frequency and choices of connectors only when they are at the initial position of sentences to eliminate other functions other than as sentence connector (see Chapter 3). Then, I report the rate of frequency of the most used choice of connectors in each corpus.

All raw frequencies are normalized by dividing them by the total number of sentences in respective corpora and multiplied by 1000 to eliminate low numbers. Then, I compare this rate of connector frequencies between the English speakers and the Asian students. Through this approach, I aim to present the contrastive list of specific connectors each group uses frequently in their writings.

English speakers vs the Asian students. The analysis of the English speakers' writing shows that they use a total of 59 different connectors at the initial position and at the rate of 145.47 incidences per 1,000 sentences. The full list of the connectors used in the English speaker's writing is available in Appendix A. On the other hand, the Asian students use 95 different connectors at the initial position and at the rate of 244.75 incidences per 1,000

44

sentences (see Appendix B). This result indicates that the Asian students use a larger number of

the connectors at the beginning position of sentence than the English speakers do.

[Table 9]

The top 10 most used connectors in the writings of the English speakers and the Asian students

Ranking	English	Asian
1	however (15.89)	so (41.40)
2	also (12.66)	but (31.19)
3	first(ly) (9.97)	and (28.25)
4	so (9.97)	however (17.23)
5	therefore (9.15)	first(ly) (11.79)
6	for example (8.08)	therefore (11.60)
7	but (7.8)	second(ly) (11.60)
8	finally (7.54)	for example (9.12)
9	second (ly) (7)	also (6.18)
10	next (6.2)	in addition (5.74)

Note: Rates of relative frequency are in parenthesis (calculated by dividing raw frequency with the total number of sentences in respective corpora and multiplied by 1,000).

Table 9 shows the top 10 most used connectors in the writings of English speakers and

Asian students in terms of choice and its rate of relative frequency. The results clearly indicate

that the English speakers and the Asian students frequently prefer to use the same connectors

even though the preference ranking is different. Both groups prefer to use the connectors *also*,

but, *first(ly)*, *for example*, *however*, *second(ly), so*, *therefore* at the initial position of sentence.

However, the connectors *finally*, *next* are only listed in the English speakers' preferred list while

and and *in addition* are listed only in that of the Asian students. Interestingly, the Asian students

frequently use the connector *and* at the initial position. On the contrary, the English speakers do

not frequently use *and* at sentence initial position. One possible explanation for infrequent use of

and at initial position is the strictures against using coordinating conjunctions at sentence initial

position. Although modern English grammars accept using coordinating conjunctions at the

beginning of independent clauses, academic writers are still most influenced by prescriptive grammar that discourages writers from starting sentences with coordinating conjunctions (Biber et al., 1999; Smith & Frawley, 1983).

A closer look into the top 10 most used connector list reveals that the Asian students use the connectors *and*, *but* and *so* considerably more than any other connectors in the list. The three connectors together account for 41.22 % of the total connectors at the beginning position of sentences. On the other hand, the English speakers appear to use the connectors diversely.

The Asian students' frequent use of *and*, *but* and *so* at the sentence-initial position raises a question on whether the English speakers use them as often as the Asian students but in a different position in a sentence. Since this analysis includes all instances regardless of the position in sentence, I use the number of words (instead of sentence) in calculating the relative frequency.

[Table 10]

The relative frequency of the connectors "so, but, and" in the writings of the English speakers and the Asian students per 1,000 words (regardless of position and function).

connector	English	Asian
and	28.16	21.76
but	3.68	6.04
so	4.33	5.61

Note: Rate of relative frequency calculated by dividing raw frequency with the total number of words in respective corpora and multiplied by 1,000

Further analysis on the connectors *and*, *but* and *so* reveals that the English speakers indeed use *and* more frequently than the Asian students (see Table 10). At the rate of 28.16 incidence per 1,000 words, the English speakers use *and* more than the Asian counterparts (21.76) per 1,000 words. Thus, we can speculate that the English speakers use the connector *and*
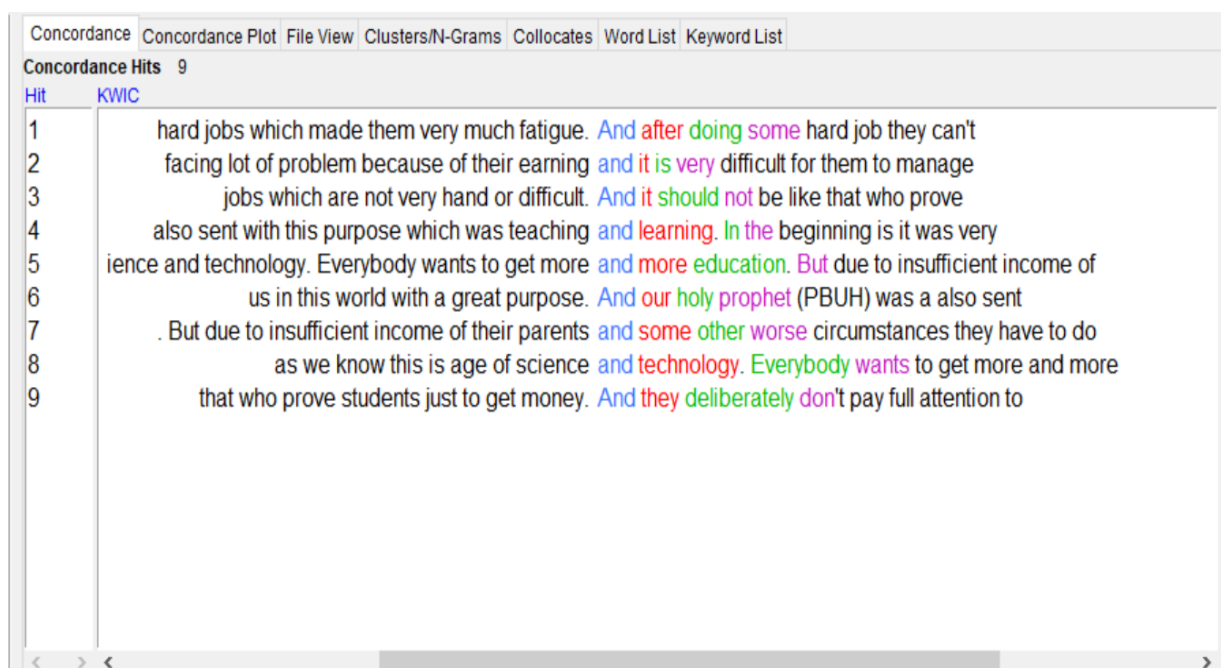
more frequently than the Asian students; however, they prefer to use it in the middle of the sentence, probably so as to combine syntactically similar elements.

Figure 5 illustrates how the English speakers make use the connector *and* in a text. It is a concordance view of the writing by an Australian student on the part-time prompt. He uses the connector *and* 11 times in a text which records the highest in the English corpus. In all of 11 incidences, he uses *and* to create a link between two ideas but never uses it once at the sentence-initial position.

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List
Concordance Hits 11

Hit	KWIC
1	I found that I was losing interest in studying and always tired. So instead of just working to study
2	rk because they wanted the money for student fees and books etc. At first, I had the same plan
3	ing lectures. Finally, I just decided to go ahead and do it and I am glad I did. So,
4	have also been studying for most of my life and felt like I needed a break. It has been
5	, I just decided to go ahead and do it and I am glad I did. So, I think that
6	needed a break. It has been great so far and I don't regret my decision one bit. My
7	ng for myself. I know that studying is worthwhile and I will finish my degree but I have also
8	k and then return to my studies feeling refreshed and like I had done something for myself. I know
9	of the money for travel, take a short break and then return to my studies feeling refreshed and I
10	me because it allowed me to take time off and travel to Japan. I know that some of my
11	father) wasn't too impressed with the idea though and tried to talk me out of it with these

[Figure 5] the use of "and" in the writing of an English speaker (EN_PTJO_004)

Figure 6 shows the use of *and* in a Pakistan student's writing on the part-time prompt. The author is a science major, and his English level is marked as advanced. His writing includes 9 incidences of the connector *and* and in five incidences he uses it at sentence-medial position like an English speaker but in three incidences he uses it at the sentence-initial position.

[Figure 6] The use of "and" in the writing of a Pakistan student (PAK_PTJO_150)

English speakers vs the Asian students' origin country. The AntConc analysis of the

English speakers' writings shows that the English speakers frequently use *however*, *also*, *first(ly),*

so, therefore, for example, *but*, *finally*, *second(ly)*, *next* by the rank of frequency. In this list, I add

the connector *and* because it is not in the preferred connector list of the English speakers but one

of the most used among the Asian students. Together, I compile the list with a total of 11

connectors and process it into the respective corpus of the Asian countries and compare the

output with that of the English speakers.

[Table 11] *Relevant value of connector frequency by the respective country, in comparison with the top 10 most frequently used connectors by the English speakers and the connector "and". The figures in parentheses are differences between the relevant values and the value in English speakers; Positive sign denotes the overuse of the connector and negative sign denotes the underuse.*

	ENG	CHI	HK	IND	JAP	KOR	PAK	PHI	SIG	TAI	THA
however	15.89	20.08 (+4.19)	27.41 (+11.52)	54.34 (+38.45)	14.52 (-1.37)	22.46 (+6.57)	1.23 (-14.66)	9.39 (+6.5)	50.79 (+34.9)	17.59 (+1.7)	12.18 (-3.71)
also	12.66	5.78 (-6.88)	10.6 (-2.06)	2.71 (-9.95)	3.33 (-9.33)	13.33 (+0.67)	1.23 (-11.43)	8.41 (-4.25)	14.23 (+1.57)	4.03 (-8.63)	3.02 (-9.64)
first(ly)	9.97	17.49 (+7.52)	6.21 (-3.78)	6.5 (-3.47)	21.02 (+11.05)	14.98 (+5.01)	2.46 (-7.51)	5.09 (-4.88)	1.31 (-8.66)	14.48 (+4.51)	6.76 (-3.21)
so	9.97	58.95 (+48.98)	13.18 (+3.21)	54.34 (+44.37)	64.66 (+54.69)	40.41 (+30.44)	38.92 (+28.95)	3.52 (-6.45)	3.72 (-6.25)	39.04 (+29.07)	29.61 (+19.64)
therefore	9.15	5.95 (-3.2)	32.6 (+23.45)	7.22 (-1.93)	11.42 (+2.27)	20.62 (+11.47)	4.92 (-4.23)	3.52 (-5.63)	17.29 (+8.14)	14.48 (+5.33)	9.78 (+0.63)
for example	8.08	6.2 (-1.88)	4.36 (-3.72)	8.3 (+0.2)	20.42 (+12.34)	11.18 (+3.1)	1.54 (-6.54)	2.74 (-5.34)	8.54 (+0.46)	10.45 (+2.37)	6.31 (-1.77)
but	7.8	28.35 (+20.55)	8.04 (+0.24)	31.23 (+23.43)	47.26 (+39.46)	37.23 (+29.43)	37.38 (+29.58)	24.46 (+16.66)	5.04 (-2.76)	30.79 (+22.99)	25.7 (+17.9)
finally	7.54	3.45 (-4.09)	2.19 (-5.35)	1.44 (-6.1)	5.22 (-2.32)	4.1 (-3.44)	0 (-7.54)	0.78 (-6.76)	0.44 (-7.1)	3.3 (-4.24)	7.74 (+0.2)
second(ly)	7	18.96 (+11.96)	3.29 (-3.71)	5.6 (-1.4)	19.43 (+12.43)	13.23 (+6.23)	0.92 (-6.08)	3.93 (-3.07)	0.44 (-6.56)	14.48 (+7.48)	6.22 (-0.78)
next	6.20	0.52 (-5.68)	0 (-6.2)	0.72 (-5.48)	1.36 (-4.84)	1.44 (-4.76)	0 (-6.2)	1.17 (-5.03)	0.66 (-5.54)	0.55 (-5.65)	1.24 (-4.96)
and	3.22	31.2 (+27.98)	7.68 (+4.46)	27.8 (+24.58)	42.04 (+38.82)	31.08 (+27.86)	19.84 (+16.62)	23.28 (+20.06)	4.6 (+1.38)	26.39 (+23.17)	29.61 (+26.39)
Total	97.48	196.93 (+99.45)	115.56 (+18.08)	200.2 (+102.72)	250.68 (+153.2)	210.06 (+112.58)	108.44 (+10.96)	86.29 (-11.19)	107.06 (+9.58)	175.58 (+78.1)	138.69 (+40.69)

Note: Rates of frequency are calculated by dividing raw frequency with the total number of sentences in respective corpora and multiplied by 1,000.

Table 11 shows the relevant value of specific connectors across the Asian students' origin country as compared to their English counterpart. Thus, a "+" signal means the specific connector is more frequent in the writings of the Asian country, whereas a "-" signal means the connector is more frequent in the English speakers' writings.

With reference to Table 9, the Asian students use more connectors at sentence-initial position when compared to the English speakers regardless of their respective countries. ESL students (Hong Kong, Pakistan, Philippine, Singapore) show a slight overuse of the connectors; however, EFL students overuse them in distinctively higher numbers. Notably, Japanese students' writings use 11 connectors considerably more often than the English speakers do, in fact, using 1.5 times more often than the English speakers.
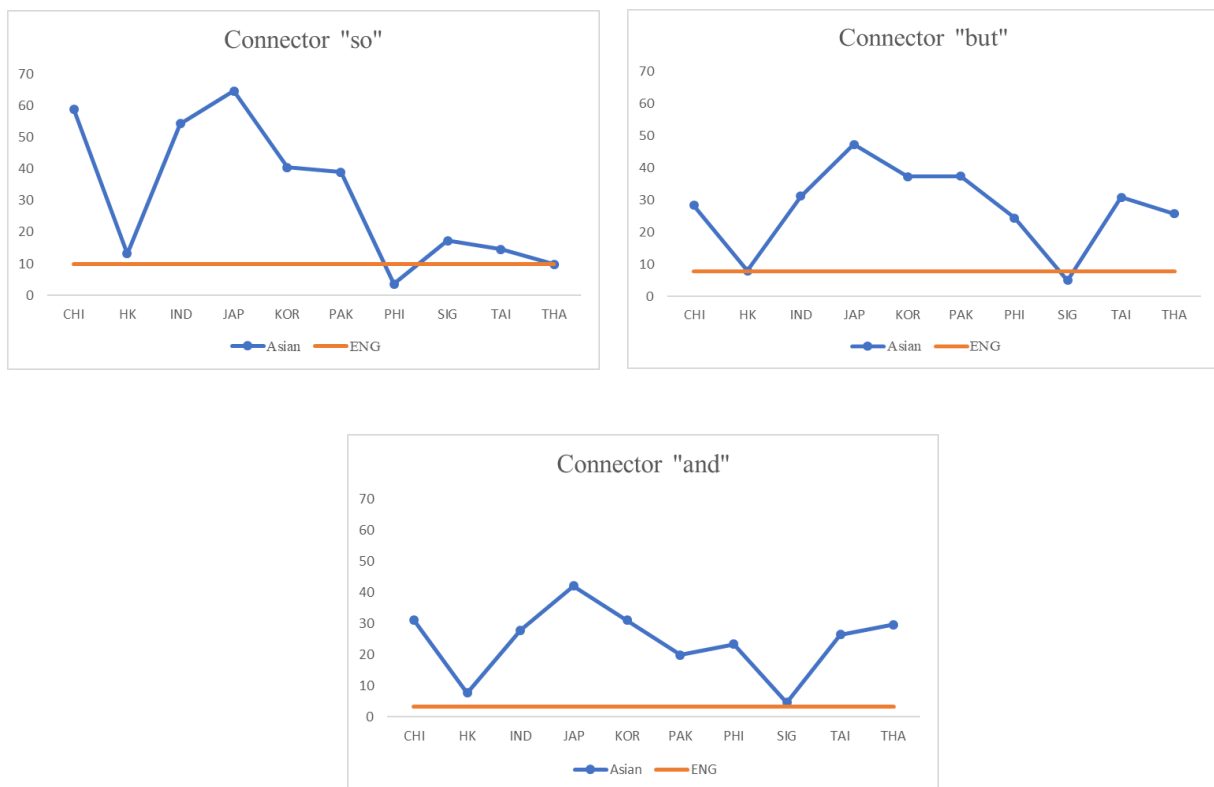
[Figure 7] The use of "so, but, and" in the writings of Asian students vs English speakers

Depending on the individual connectors, the Asian students show a varied rate of frequency. However, they consistently overuse the connectors *and*, *but*, *so* except for Hong Kong, Philippine and Singapore students (see Figure 7). Japanese students use three connectors, the most with the greatest difference from the norm of English speakers. They use connector *so* in 54.69 more incidences per 1,000 sentences than English speakers, using it about six times as often as the English speakers, the connector *and* about 41 times, and *but* about 3 times as often as the English speakers. Although Hong Kong students do not show differences in the use of the connectors *and*, *but*, *so,* they use considerably more often the connectors *however* and *therefore*. Similarly, Singapore students use *however* considerably more than the English speakers.

In contrast to the Asian students' tendency to overuse many of the connectors when compared to English speakers, only a few connectors are underused by the Asian students compared to the English speakers. The connector *next* has 0.7 instances on the average in the Asian students' writing while the English speakers' writings have 6.2 instances.

Taken together, the Asian students appear to greatly overuse many of the connectors in the selected list across the countries. Especially the connectors *so*, *and*, *but* are considerably overused by the Asian students from many countries, using them as much as 41 times more than the English speakers per 1,000 sentences. Japanese students' overusing tendency is worth noting, too. On the other hand, only a few connectors are underused by the Asian students as compared to the English speakers and their level of underuse is noticeably lower than those for overuse.

English speakers vs the Asian students' English proficiency level. In this stage of the analysis, I use the compiled list of 11 connectors, comprising the top 10 most used connectors by the English speakers and the connector *and* since it is frequently used by the Asian students. I process this list in Asian students' writings depending on their English proficiency

level using AntConc.

[Table 12]

Relative frequency of the 11 connectors in the Asian students' writing by their English proficiency level (+/- show the difference between the value of the English speakers and the value in respective country; a positive value denotes overuse and a negative value denotes underuse.)

	ENG	Beginner	Int-Low	Int-High	Advanced
however	15.89	16.10 (+0.21)	12.70 (-3.19)	17.58 (+1.69)	38.27 (+22.38)
also	12.66	5.46 (-7.2)	4.41 (-8.25)	7.32 (-5.34)	11.18 (-1.48)
first(ly)	13.64	14.40 (+0.76)	12.91 (-0.73)	9.27 (-4.37)	11.50 (-2.14)
so	10.80	60.38 (+49.58)	47.37 (+36.57)	33.22 (+22.42)	22.99 (+12.99)
therefore	10.22	14.40 (+4.18)	9.61 (-0.61)	11.25 (+1.03)	19.37 (+9.15)
for example	8.08	10.71 (+2.63)	10.17 (+2.09)	7.28 (-0.8)	11.18 (+3.1)
but	7.81	44.90 (+37.09)	32.41 (+24.6)	28.03 (+20.22)	19.84 (+12.03)
finally	7.54	5.66 (-1.88)	4.58 (-2.96)	2.25 (-5.29)	2.68 (-4.86)
second(ly)	7.00	13.65 (+6.65)	12.70 (+5.7)	7.89 (+0.89)	10.55 (+3.55)
next	6.20	1.09 (-5.11)	0.94 (-5.26)	0.88 (-5.32)	0.79 (-5.41)
and	3.23	48.31 (+45.08)	30.23 (+27)	23.53 (+20.3)	13.07 (+9.84)
Total	103.07	235.06 (+131.99)	178.03 (+74.96)	148.5 (+45.43)	161.42 (+58.35)

Note: Rate of frequency are calculated by dividing raw frequency with the total number of sentences in respective corpora and multiplied by 1,000.

Table 12 shows the comparative analysis of 11 connectors across the four English proficiency levels (beginner, intermediate-low, intermediate-high, advanced). When analyzing all the connectors together, it appears that the Asian students gradually use fewer connectors at the sentence-initial position through the Intermediate high level. Then, the total number of the connectors increases at the advanced level, particularly as a result of the overused connector *however* (see Figure 8).
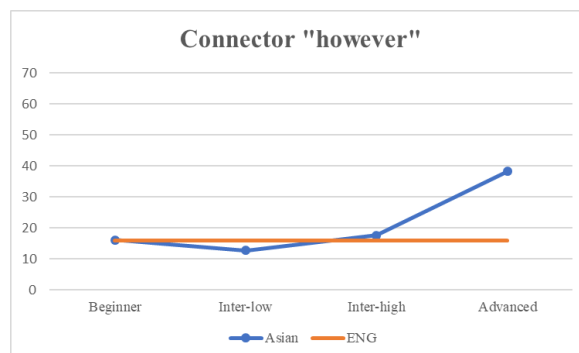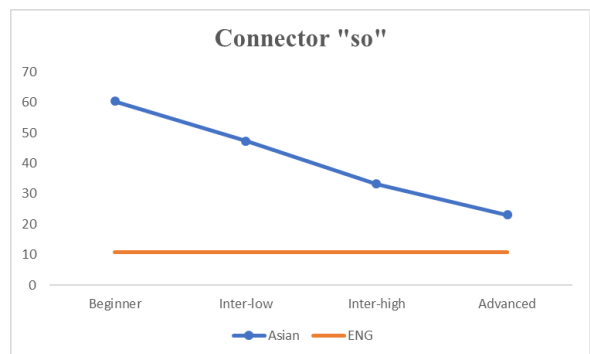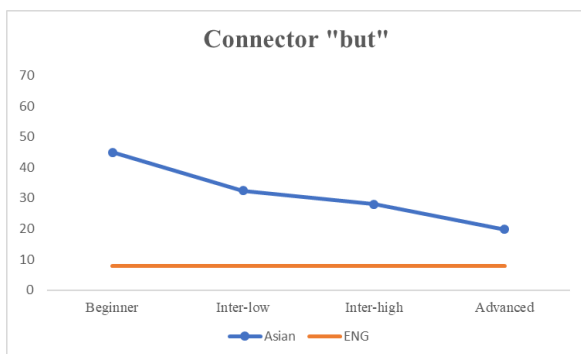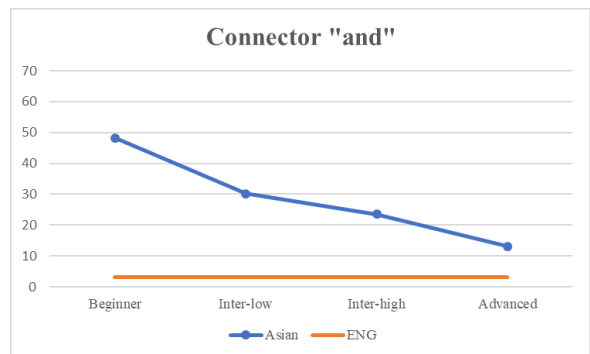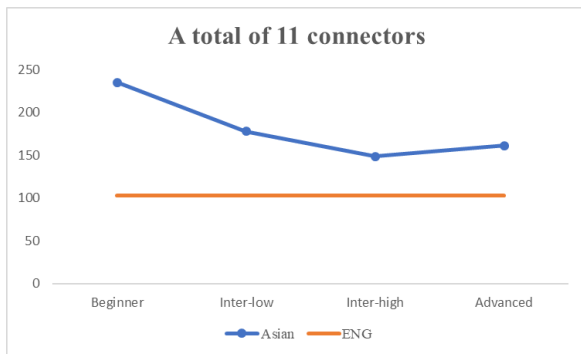
[Figure 8] Connectors in Asian students' writing by proficiency level vs English speakers

In the beginners, intermediate low and intermediate high levels, the connectors *so, and*, and *but* are the top 3 most overused at the initial position, as compared to the English counterparts. Their rate of occurrence tends to follow the model of the English speakers as the English proficiency level increases (see Figure 8). Specifically, the connector *so* is used in the beginning level at the rate of 60.38 and intermediate low level at the rate of 47.36, which is about five times more than the English speakers (9.09). Then, in the intermediate high

level, *so* is used about three times (33.22) and in the advanced level at about two times (22.99). As such, the Asian students appear to acquire a better understanding of the use of *so*, *and*, *but* and don't use them as often at the initial position as their English improves; however, even in the advance level they still show   quite a large number of the three connectors at the initial position (+12.99).

While the connector *so* is most overused in beginner and intermediate low and high level, the advanced level students use the connector *however* most frequently (see Figure 8). At the rate of 38.27 occurrences per 1,000 sentences, the connector *however* is used about three times more than the intermediate low level (12.70), twice as much as the beginner level (22.13), the intermediate high level (17.58), and the English speakers (15.89).

Only a few connectors are underused by Asian students compared to the English speakers (see Table 12). Across the proficiency levels, the connectors *also*, *finally*, *next* occur less often; however, their level of underuse is much lower than those for overuse.

Taken together, the Asian students in four English proficiency levels appear to gradually follow the model of the English speakers in the use of many connectors. However, their use of certain connectors like *so*, *and*, *but* are considerably overused across all levels. Additionally, the advanced level students prefer to use the connector *however* when compared to the Asian students in other levels and to the English speakers.

Previous analyses on the use of the popular connectors *and*, *but*, *so* reveal that the English speakers actually use *and* more than the Asian students (see Table 10), but not at the sentence-initial position. Thus, the decreasing use of the three popular connectors raises a question whether the Asian students indeed learn to place those connectors in non-initial positions in a sentence or simply learn not to use it at the initial position as their English level increases. Since this analysis includes all instances regardless of the position in sentence, I use the number of words (instead of sentences) as a base unit in calculating the relative

frequency.

[Table 13]

The relative rate of frequency of the connectors "so, but, and" in the writings of the English speakers and the Asian students in four levels per 1,000 words when they are not at initial position.

connector	English	Beginner	Int-Low	Int-High	Advanced
and	28.02	19.98	19.31	20.74	19.55
but	3.36	4.50	4.35	3.86	3.35
so	3.92	3.31	3.32	2.82	2.26

Note: Rate of frequency is calculated by dividing raw frequency with the total number of words in respective corpora and multiplied by 1,000.

Table 13 shows that the relative frequency of the three connectors *and*, *but*, *so* when they are not at the sentence-initial position. The Asians students across the four levels use the connector *and* considerably less at non-initial positions than the English speakers. Moreover, the difference between the levels is limited. Thus, we can speculate that the Asian students appear to fail grasping the use of *and* in a non-initial position even as their English progresses. Instead, they simply refrain from using it. For the connectors *but* and *so*, the Asian students in all four levels show not much difference from the norm of English speakers.

Collectively, while the Asians students learn to use the three most frequently used connectors (*and*, *but*, *so*) less often at the sentence-initial position as their English improves, it appears that they do not use the connector *and* as much as their English counterpart at the sentence-medial position across the levels. However, they use the connectors *but* and *so* in a non-initial position as much as the English speakers regardless of their English proficiency level.

Chapter 5. Discussion and Conclusion

In this study, I examined the writings of the Asians students from 10 different countries (Hong Kong, Pakistan, the Philippines, Singapore, China, Indonesia, Japan, Korea, Taiwan, and Thailand) and compared their writings with those of the English speakers on the use of connectors. Specifically, I explored whether the Asian students use the connectors similarly or differently compared to English speakers in terms of frequency and choice, depending on their origin countries and English proficiency levels. By highlighting their linguistic pattern, the study aims to heighten the awareness of connector use and ultimately help the Asian students to compose more native-like writings.

The corpus of the current study was culled from the written essay module of the ICNALE (the International Corpus Network of Asian Learners of English). All the writings were processed through two computation tools: Coh-Metrix and AntConc. Coh-Metrix was used to broadly assess the frequency of all the connectors used in the respective corpus and of seven subcategories (casual, logic, contrastive, temporal, additive, positive, and negative connectors). In analyzing data generated by Coh-Metrix, I conduct MANOVA (Multivariate Analysis of Variance) in order to determine whether the difference in the descriptive data is statistically significant, followed by Tukey post-hoc test. AntConc analysis was conducted to evaluate the preference and frequency of specific connectors that occurred in the respective corpus.

I will discuss the findings by addressing the first research question: *Is there a similarity in the use of connectors between the writing of college students in 10 Asian countries and native speakers (NS) of English, in terms of frequency and choice of connector? If there is, how do they employ the connectors similarly?* Only one index in Coh-Metrix showed a similarity between the Asian students and the English speakers in their written production, which is contrastive (e.g. *although*, *whereas*) connector. In the AntConc

analysis on the choice of connectors, the English speakers and the Asian students share a great deal of overlap in the list of high-frequency connectors at the sentence-initial position such as *also*, *but*, *first(ly)*, *for example*, *however*, *second(ly), so* and *therefore*. As such, the English speakers and Asian students appear to use a similar number of contrastive connectors and prefer to use a similar set of the connectors at sentence-initial position.

Addressing my second research question, *Is there a difference in the use of connectors between the writing of college students in 10 Asian countries and native speakers of English in terms of frequency and choice of connectors? If so, how do they employ the connectors differently?* As expected, the results found more differences than similarities on the connector use between the Asian students and English speakers. Firstly, Coh-Metrix analysis supported that there was no statistical difference between the Asian students and the English speakers in the number of the all types of connectors per 1,000 words. However, the two groups preferred different types of connectors in their writings. Overall, the Asian students preferred logical (e.g. *and*, *if-then*) and positive connectors (e.g. *also*, *moreover*) more than other types of connector. In contrast, the English speakers chose additive and positive connectors as the top preferred connector types than the others. As such, in creating cohesive links between related ideas, the Asian students heavily relied on logical and positive connectors whereas the English native speakers preferred positive and additive connectors.

Comparison between the Asian students and the English speakers revealed that the Asian students used significantly more logical and positive connectors than their English counterparts. On the other hand, they used a significantly fewer number of causal, additive and negative connectors. Accordingly, it might be beneficial to the Asian students if they are encouraged to recognize when to use causal, additive, and negative connectors in building a bond between two propositions. In the attribute of contrastive connectors, there is no significant difference between the Asian students and the English speakers.

In addition to the shared list of the frequently used connectors, the Asians students used the connector *and* frequently at sentence-initial position. They also showed a heavy dependence on the three connectors *and, so* and *but* in joining two sentences. On the other hand, the English speakers used the connecters more diversely.

While both groups used a similar set of connectors at the sentence-initial position, they positioned certain connectors differently within a sentence. The English speakers preferred *and, so* and *but* in the medial position of a sentence; however, the Asian students preferred to use them greatly at the sentence-initial position, which supports the findings of Narita, Sato and Sugiura (2004). This tendency could be attributed to the fact that the EFL students might be more comfortable in using cohesive devices at sentence-initial position to express an explicit bond between two ideas (Rutherford, 1987). Another plausible explanation is that the Asian students do not have a sufficient linguistic knowledge on the flexible position of connectors (Narita, Sato & Sugiura, 2004).

In reference to my third research question, *is there a difference in the use of connectors between the writings of college students within 10 Asian countries depending their English proficiency level and their nationalities? If so, how do they employ the connectors similarly and differently across English level and nationality?* I will first discuss the similarity and difference between the English speakers and the Asian students depending on their nationalities. When compared to the English speakers, all of the Asian students across their countries showed a strong evidence of the overuse of logical and positive connectors and underuse of additive and negative connectors regardless of their origin countries. However, the analysis on their use of causal, contrastive and temporal connectors yielded a mixed result depending on their countries.

When combining the frequencies of the selected 11 connectors (*and, however, also, first(ly), so, therefore, for example, but, finally, second(ly), next*) at sentence-initial position,

we noticed that the Asian students in all countries used them more than the English speakers. However, ESL countries (Hong Kong, Pakistan, Philippine, Singapore) used the 11 connectors with only slight differences to the English speakers. On the other hand, EFL countries (Chinese, Indonesia, Japanese, Korean, Taiwan, Thailand) substantially overused those connectors at sentence-initial position compared to the English speakers.

Regarding the frequency of the individual connectors, the Asian students showed variances in preference depending on their origin countries; however, a great majority of the them shared an overreliance on the three connectors *and*, *but*, and *so* to a higher extent, except for Hong Kong, Philippine and Singapore students. Noticeably, Japanese students overused the three connectors and showed the greatest difference from the English speakers. On the other hand, Hong Kong and Singapore students chose *therefore* and *however* respectively as their top preferred connector.

Finally, I will discuss the findings on the connector use between the English speakers and the Asian students depending on their four English proficiency levels (beginner, intermediate low, intermediate high, advanced). In the Coh-Metrix analysis on overall use of connectors, the Asians students show a decrease in the total number of connectors used as they develop English proficiency; however, the difference was not statistically significant.

Detailed analyses on seven classifications of the connectors revealed that in temporal types of connectors, the Asian students gradually followed the model of the English speakers, using more as their English progressed. On the other hand, the logic and positive types of connectors were significantly overused in their writings across the proficiency levels. In addition, the Asian students underused additive and negative connectors across the proficiency levels.

In the AntConc analysis on the selected 11 connectors, the Asian students in all proficiency levels used substantially more connectors at sentence-initial position than their

English counterparts. It is encouraging that they gradually used fewer connectors as their English progressed. However, such decreasing tendency was discontinued when they reached an intermediate-high level. Then, their total frequency of the 11 connectors increased in the advanced level, particularly due to their sudden increase in the use of *however*.

The analysis on the individual connectors revealed that the connector *so* was used most frequently in the beginner, intermediate-low, and intermediate-high levels, followed by *and* and *but*. In the advanced level, the connector *however* was most preferred. As compared to English speakers, the Asian students showed a strong preference on three connectors *and*, *but* and *so*, regardless of their English levels. However, as their English improved, their preference on the three connectors at sentence-initial position showed a gradual decrease.

A closer look at the connectors *and*, *but* and *so* on their positions in a sentence shows the Asians students used the connector *and* less at sentence-medial position across the levels, when compared to the English speakers. On the other hand, at sentence-medial position, the Asian students showed no difference in the number of the connectors *but* and *so*. Conversely, as their English improved, the Asian students appear to follow the model of the English speakers using *and* less at sentence-initial position. However, they do not seem to gain the linguistic knowledge to use *and* in the middle of sentence as many as the English speakers (possibly to form compound sentences), even though their English improved.

Taken as a whole, the current study demonstrated that the Asian students and English speakers shared common features on the connector usage such as the set of preferred connectors at sentence-initial position. However, their differences were far greater. The Asian students' writings were characterized with a greater number of logical and positive connectors and a fewer number of additive and negative connectors compared to the English speakers. Moreover, they showed an overreliance on certain connectors like *and*, *but*, and *so* at the beginning of sentence while the English speakers preferred to use them in the middle of

sentences.

The findings suggest that, in order to compose more native-like writings, the Asian students need to focus more on using additive and negative connectors and decrease the dependence on logical and positive connectors in expressing relations between syntactically (or semantically) related ideas. Additionally, they need to use the connectors diversely, and learn to position them more in the middle of sentence, rather than the beginning of sentence.

Limitation of the study

Although the findings of this study shed light on the linguistic features of writings between the Asian students and the English speakers, some limitations should be noted. Firstly, the study does not encompass the writings of a wide variety of topics nor genres. It includes the argumentative writings only on two topics: the part-time and smoking prompts; thus, it requires some caution when generalizing the findings into different genres or topics.

While the study has looked closely into the three connectors (*and, but, so*) in different positions within a sentence, it primarily concerns the use of specific connectors at sentence-initial position. Therefore, another limitation is that the current study only examines a small list of connectors at sentence-medial position. A larger list of connector analysis in sentence-medial position will help further identify the linguistic features on the connecter use between the Asian students and the English speakers.

Future study

The current study only begins to reveal the linguistic features of connector use in the Asian students compared to the English speakers. Further study is necessary to examine the practice of connectors with a large scope of topics and different genres such as narratives, not being limited to two topics and one genre as in the current study. Thus, we will have a more comprehensive understanding of the Asian students' language use of connectors, which is an important cohesive device in the quality of written production.

Additionally, the current study creates a basis for future qualitative analysis on the connectors' positions and their function within a sentence. Such research could and should examine how the Asian students position individual connectors and, further, whether they prefer to use them in connecting nouns, phrases or other grammatical functions in a sentence compared to the English speakers' preferences in writing.

Conclusion

While there is still much work to be done, the findings of the study give a glimpse at the Asian students' linguistic choices on connectors across their countries and proficiency levels. A key finding suggests that the Asian students use fewer contrastive and negative connectors, prefer certain connectors at sentence-initial position but not as often in sentence-medial position compared to the English speakers. It is also encouraging to observe that Asian students appear to follow the norm of the English speakers in the use of certain connector type as their English progresses.

The findings of this study contribute to the current understanding of how the Asian students distribute connectors in their writings and further how the English speakers as a norm use them in comparison. Thus, the results offer a helpful instructional resource to ESL writing educators and textbook designers as to what area they need to focus on when teaching connector usage to the Asian students. Furthermore, for the Asian students themselves, the findings will give an opportunity to reflect on their writing habits and raise an awareness of connector usage, ultimately to help them increase their mastery of cohesive devices.

References

Al-Badi, I (2015). Academic writing difficulties of ESL learners. WEI International
Academic Conference Proceedings. *8*, 63-76.

Anthony, L. (2016). *AntConc*: Design and development of a freeware corpus analysis toolkit
for the technical writing classroom. IEEE International Professional Communication
Conference Proceedings, *2* (2), 99-127.

Anthony, L. (2018). AntConc (Version 3.5.7) [Computer Software]. Tokyo, Japan: Waseda
University. Available from http://www.laurenceanthony.net/software

Azadnia, M., Lotfi. A. R. & Biria, R. (2019). A study of syntactiv complexity via Cho-
Metrix: Similarities and differences of Ph.D. dissertation written by Iranian University
Students and English native speakers. *Research in English Language Pedagogy*. *7* (2).
232-254.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). The Longman
Grammar of Spoken and Written English. London: Longman.

Blakemore, D. (2002) Relevance and Linguistic Meaning: The Semantics and Pragmatics of
Discourse Markers, Cambridge University Press.

Bolton, K., Nelson,m 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*(2), 165-182.

Brezina, V., McEnery, T., &Wattam, S. (2015). Collocations in context: A new perspective on
collocation networks. *International Journal of Corpus Linguistics, 20*(2), 139-173.

Casanave, C. P., & Hubbard, P. (1992). The Writing Assignments and Writing Problems of
Doctoral Students: Faculty Perceptions, Pedagogical Issues, and Needed Research.
English for Specific Purposes, 11(1), 33-49.

Chang, C. & Kuo, C. (2011). A corpus-based approach to online materials development for writing research articles. *English for Specific Purposes*, *30*, 222-234.

Chanyoo, N. (2018). Cohesive Devices and Academic Writing Quality of Thai Undergraduate Student. *Journal of Language Teaching and Research*, 9(*5*). 994-1001.

Chou, L. (2011). An Investigation of Taiwanese Doctoral Students' Academic Writing at a U.S. University. *Higher Education Studies*, *1*(2), 47-60.

Conrad, S. (1996). Investigating academic texts with corpus-based techniques: an example from biology. *Linguistics and Education*, *8*, 299-326.

Cowan, R., (2008). The teacher's grammar of English. Cambridge: Cambridge University Press.

Crewe, W. (1990). The illogic of logical connectives. *ELT Journal*, *44* (4), 316–325.

Crossley, S., Kyle, K., & McNamara, D. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, 32, 1-16.

Crossley, S. A., & McNamara, D. S. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. In Proceedings of the 32nd annual conference of the Cognitive Science Society (pp. 984–989). Austin, TX: Cognitive Science Society.

Crossley, S. A., & McNamara, D. S. (2011). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life-Long Learning*, 21(2/3), 170–191

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, 35, 115–135.

Crossley, S., Salsbury, T., & McNamara, D. (2009). Measuring L2 lexical growth using hypernymic relationships. *Language Learning*, 59(2), 307–334.

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, 10(1), 5–43.

Erasanli, C. Y. (2015). Insights from a Learner Corpus as Opposed to a Native Corpus about Cohesive Devices in an Academic Writing Context. *Universal Journal of Educational Research, 3*(12), 1049-1053.

Ferris, J. (1992). A computer text analysis of four cohesion devices in English discourse by native and nonnative writers. Journal of Second Language Writing 1.2, 79-107.

Ferris, D. & Hedgcock, J. (2005). Teaching ESL composition. New Jersey, Lawrence Erlbaum Associates.

Flowerdew, L. (2015). Using corpus-based research and online academic corpora to inform writing of the discussion of a thesis. *Journal of English for Academic Purposes*, 20, 58-68.

Halliday, M. A. K., & Hasan, R. (1976). Cohesion in English. London: Longman.

Halliday, M., & Hasan, R. (1985). Language, context, and text: Aspects of language in a social semiotic perspective. Deakin: Deakin University.

Hinkel, E. (2002). Second language writer's text: Linguistic and rhetorical features. New Jersey:

Lawrence Erlbaum Associate.

Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian Learners of English. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world, 1,* 91-118. Kobe, Japan: Kobe University.

John, T. (1994). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In Terence Odlin (Ed.), *Perspectives on pedagogical grammar*. Cambridge: Cambridge University Press.

Liu, M., & Braine, G. (2005). Cohesive Features in argumentative writing produced by Chinese undergraduates, *System*, 33(4), 623-636.

Liu, G. (2013). On the use of linking adverbials by Chinese College English learners. *Journal of Language Teaching and Research*, 4 (*1*), 149-155.

Martinez, R. (2018). "Specially in the last years…": Evidence of ELF and non-native English forms in international journals. *Journal of English for Academic Purposes*, 33, 40-52

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57–86.

McCarthy, P. M., Lehenbauer, B. M, Hall, C., Duran, N. D., Fujiwara, Y. & McNamara, D.S. (2007). A Coh-Metrix analysis of discourse variation in the texts of Japanese, American, and British Scientists. *Foreign Languages for Specific Purposes*, *6*, 46-77.

Min, H.S., & Mccarthy, P. M. (2013). Contrastive corpus analysis on the writing of American and Korean scientists using Gramulator. *Language Research*, 49 (*3*), 681-702.

Narita, M., Sato, C., & Sugiura, M. (2004). Connector usage in the English essay writing of Japanese EFL learners. Paper presented at the Fourth International Conference on

Language Resources and Evaluation.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). A Comprehensive Grammar of the English Language. London: Longman Group Limited.

Ravid, D. D. & Shyldkrot, H. B. (Eds.). (2005). *Language Development*: *Essays in Honor of Ruth A. Berman*, New York, NY: Kluwer Academic Publishers.

Reid, J. (1992). A computer text analysis of four cohesion devices in English discourse by native and nonnative writers. Journal of Second Language Writing 1.2, 79-107.

Römer, U. & Wulff, S. (2010). Applying corpus methods to writing research: Explorations of MICUSP. *Journal of Writing Research*, *2*(2), 99-127.

Rutherford, W. (1987). Second Language Grammar: Learning and Teaching. Longman.

Shea. M. (2009). A Corpus-Based Study of Adverbial Connectors in Learner Text. *MSU Working Papers in SLS, 1* (1).

Student and Exchange Visitor Program. (2017). SEVIS by the numbers: Biannual report on international student trends. Retrieved from https://www.ice.gov/doclib/sevis/pdf/byTheNumbersJun2017.pdf

Smith, R. & Frawley, W. (1983). Conjunctive cohesion in four English genres. Text 3, 347–374.

Taguchi, N., Crawford, W. & Wetzel, D. Z. (2013). What Linguistic Features Are Indicative of Writing Quality? A Case of Argumentative Essays in a College Composition Program. *TESOL Quarterly*, 47 (2).

Thurstun, J., & Candlin, C. N. (1998). Concordancing and the Teaching of the Vocabulary of

Academic English. *English for Specific Purposes,17*.

Yang, W., & Sun, Y. (2012). The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. Linguistics and Education, 23(1), 31–48.

Ye, D., & Wang, D. (2013). Language Varieties of the Abstracts in Journal Articles Written by Chinese and American Scholars: A Contrastive Corpus Analysis using the Gramulator. *International Journal of Applied Linguistics & English Literature*, *2*(2).

Ye, D. (2013). A Coh-metrix analysis of language varieties between the journal articles of Chinese and American scientists. *International Journal of English Linguistics, 3*(4), 63-70.

Yunxia, S., Min, Y., & Zhuo, S. (2009). An empirical study on a computer-based corpus approach to English vocabulary teaching and learning. Paper presented at the 2nd International Conference on Computer Science and Information Technology, Beijing, China

Appendix A

The sentence-initial connectors in English speakers' corpus

(The figures in parentheses are raw frequencies, Total number of sentences = 3,712)

Connectors	Frequency (x1000)
however	15.89 (59)
also	12.66 (47)
first(ly), so	9.97(37)
therefore	9.15(34)
for example	8.08(30)
but	7.8(29)
finally	7.54(28)
second(ly)	7(26)
next	6.20(23)
further(more)	5.12(19)
in addition	3.50(13)
and	3.23(12)
Third(ly)	2.96(11)
last(ly)	2.69(10)
in conclusion	2.42(9)
on the other hand	2.16(8)
as a result	1.86(7)
at least, in the same way, indeed, then, thus	1.89(5)
additionally, all in all, first of all, for this reason, nonetheless, or, overall	1.08(4)
consequently, conversely, in fact, in short, instead, what's more, yet	0.81(3)
accordingly, besides, despite this, for instance, otherwise	0.54(2)
Actually, anyway, by the same token, for a start, for one thing, in any case, in other words, in spite of, in summary, moreover, neither, nevertheless, nor, on the contrary, rather, to begin with, to summarize	0.27(1)
Total	145.47 (540)

Appendix B

The sentence-initial connectors in Asian students' corpus (The figures in parentheses are raw frequencies, Total number of sentences = 73,002)

Connectors	Frequency (x1000)
so	14.4(3022)
but	31.19(2277)
and	28.25(2062)
however	17.23(1258)
first(ly)	11.79(861)
therefore	11.6(847)
second(ly)	10.99(802)
for example	9.21(672)
also	6.18(451)
in addition	5.74(419)
moreover	5.41(395)
then	5.41(395)
third	4.59(335)
besides	4.27(312)
finally	3.67(268)
first of all	3.60(263)
thus	3.45(252)
in conclusion	3.41(249)
on the other hand	2.81(205)
in fact	2.67(195)
further(more)	2.36(172)
last	2.01(147)
hence	1.9(139)
as a result	1.88(137)
actually	1.71(125)
for instance	1.68(123)
though	1.51(110)
or	1.47(110)
in a word	1.23(90)
all in all	0.96(70)
next	0.92(67)
in other word	0.85(62)
instead, to sum up	0.81(59)
in short	0.79(58)
what's more	0.7(51)
nevertheless	0.67(49)
otherwise, that is	0.56(41)
by the way	0.52(38)
to begin with	0.51(37)

The sentence-initial connectors in Asian students' corpus (The figures in parentheses are raw frequencies, Total number of sentences = 73,002)

Connectors	Frequency (x1000)
anyway	0.48(35)
indeed	0.47(34)
additionally, yet	0.44(32)
at least	0.42(31)
for this reason	0.4(29)
that is to say	0.34(25)
consequently	0.32(23)
on the contrary, still	0.3(22)
in spite of, in the first place	0.27(20)
in that case	0.26(19)
for one thing	0.23(17)
as a matter of fact, in summary, in the second place, rather	0.21(15)
accordingly, overall	0.18(13)
similarly	0.14(10)
as a consequently, in sum	0.12(9)
in contrast, likewise, nonetheless	0.11(8)
anyhow	0.1(7)
alternatively, in reality	0.08(6)
arising out of, in the same way, to tell the truth	0.07(5)
conversely, for another thing, for this purpose, on top of that, to summarize	0.05(4)
in this respect, neither	0.04(3)
aside from this, at this point, from now on, in any case, in consequence, last of all, namely, nor, to conclude with	0.03(2)
As it happens, by contrast, despite this, hitherto, in either case, on account of	0.01(1)
Total	244.75(17867)