

# A Corpus-based Analysis of TESOL EFL Students' Use of Logical Connectors in Spoken English

Hao Wu

Faculty of Education, The University of Hong Kong, Hong Kong SAR, China

**Abstract**—In the past several decades, the compilation of learner corpora and the application of corpus linguistics have been extensively employed to improve learners' use of logical connectors. However, the use of logical connectors in EFL learners' spoken discourse remains under-researched. To investigate this field, the researcher built an EFL TESOL student spoken English corpus consisting of 27 spoken English samples of 12,241 words in total. Then, this study adopts corpus-based contrastive analysis and computer-aided error analysis to compare the tokens and the frequencies of the logical connectors with those in the native spoken English corpus of MICASE. Finally, underuse, overuse, and misuse in the TESOL student corpus were exemplified and explained.

Findings reveal that the TESOL students tended to use a smaller set of logical connectors but used them more frequently than the English native speakers. Additive coordinating conjunctions such as *and*, *so*, and *but* were the most overused logical connectors. Moreover, the underuse of *if*, *when*, *so that*, and *though* shows that adverbial clauses were less frequently employed in their spoken discourse. A detailed explanation and pedagogical implications are also listed to help learners understand how to contextualize logical connectors at both syntactic and discourse level.

**Index Terms**—corpus linguistics, spoken English, logical connectors

## I. INTRODUCTION

Coherence and clarity are characteristics of comprehensible and effective communication in both written and spoken English, which can be realized by accurate and appropriate use of logical connectors of various types. According to Chafe (1994), logical connectors coordinate chunks of text and organize information flow in discourse. In other words, logical connectors help listeners comprehend and make sense of discourse. As Altenberg and Tapper (1998) mentioned, each language has a unique set of logical connectors due to the uniqueness of language and culture. Some languages such as English need overt logical connectors to indicate cohesion and coherence in the written and spoken discourse while others such as Chinese do not. Therefore, learners' native languages have the potential to influence their use of logical connectors in their second/foreign languages.

Due to the importance of the EFL learners' use of logical connectors, it can be noticed that applied linguists all around the world have participated in relevant research from the perspective of syntactical positions, frequencies, and range. In terms of syntactical positions, Green, Christopher, and Mei (2000) suggested that in English for Academic Purposes writing, Chinese EFL learners displayed their preference to place logical connectors in sentence-initial position when they wanted to introduce new information to the audience. This is consistent with the findings by Kusuyama (2003), who reported that compared with native English writers, Japanese ESL writers tended to overuse, misuse, and underuse logical connectors in sentence-initial position at the university level. Identically, Lee (2004) found that Korean learners also preferred placing them in the sentence-initial position. For the aspect of frequencies and range, based on the International Corpus of English in Hong Kong (ICE-HK) and in Great Britain (ICE-GB), Bolton, Nelson, and Hung (2002) focused on logical connector usage in Hong Kong and British university students' writing and found that both native speakers and non-native speakers of English tended to overuse various kinds of logical connectors, but they did not significantly underuse any of them. However, Lee (2004) compiled a Korean EFL learner corpus and found that Korean EFL learners obviously overused and underused some logical connectors and tended to repeat the use of logical connectors of the same types (i.e. enumeration and addition conjunctive adverbials and contrast and concession conjunctive adverbials).

In addition to the research studies above on learners' performance in written discourse, there are also ones on spoken discourse. For example, Schleppegrell (1996) demonstrated how ESL used *because* in spoken and written discourse. He realized that ESL writers used *because* to display connections between parts of discourse, to introduce segments, and to suggest the knowledge base for their declarations while in their spoken English, *because* was used quite the opposite. However, here the use of logical connector *because* was studied by case instead of an overall observation of range and frequencies in a corpus. Later, Fung and Carter (2007) discussed native speakers' and learners' use of discourse markers in spoken English in pedagogical settings based on a sub-corpus of CANCODE. They found Hong Kong

learners employed referentially functional discourse markers (e.g. *and*, *but*, *because*, and *OK*) more freely while native speakers tended to use a wider range DMs. However, the range of logical connectors is far wider than that of DM according to Halliday (1976). Therefore, a probe into logical connector frequencies and range may help teachers of English understand the influence of learners' L1 on the L2 acquisition, develop more effective teaching strategies, and fill the research gap of the lack of investigation of the frequencies and range of EFL learners' use of logical connectors. Therefore, the present study explores this very topic of EFL learners' application of logical connectors in their spoken discourse, which was examined and analyzed with the help of corpora. The present study first compiled an EFL learner corpus, the contributors of which were studying in Master of Arts in Teaching English to Speakers of Other Languages (MA(TESOL)) at a university in Hong Kong. Then, the study compared it with the well-built spoken English corpus of Michigan Corpus of Academic Spoken English (MICASE) to sort out some key features of the two groups.

## II. THEORETICAL BACKGROUND

Corpus is a useful tool to diagnose the patterns and analyze the use and the contexts of logical connectors. This study adopts a corpus-based approach as the main research method. Based on previous research studies on logical connectors in EFL learners' written discourse, this section first introduces how this study categorizes logical connectors and how "use" is defined and measured. Then, this section reviews the compilation approaches of EFL learner corpora and the methodology of corpus-based analysis.

### A. Classification of "Logical Connectors"

Logical connectors can be used to join or connect two thoughts that are connected by different kinds of relationship. According to Sorensen (1997), four kinds of function words can be logical connectors: (1) subordinating conjunctions (e.g. *before*, *after*, *since*), (2) coordinating conjunctions (e.g. *for*, *and*, *nor*, *but*, *or*, *yet*, *so*), (3) prepositions (e.g. *due to*, *because of*), and (4) transitions and conjunctive adverbs (e.g. *however*, *furthermore*, *nevertheless*). These relationships can be further categorized into: sequential (time), reason and purpose, adversative (opposition and/or unexpected result), and condition (Sorensen, 1997). Halliday (1976, pp. 242-243) also proposed a systematic and well-categorized summary of connectors of various logical relationships within written and spoken discourse. He categorized logical connectors into additive, adversative, causal, and temporal. Tailoring to EFL learners' use of logical connectors in spoken discourse, this study adopts Sorensen's classification of part of speech and Halliday's classification of relationship as the framework and provides some examples of classified logical connectors in Table 1.

TABLE I.  
SUMMARY OF LOGICAL CONNECTORS BASED ON SORENSEN (1997) AND HALLIDAY (1976)

Part of speech / Relationship	Subordinating conjunctions	Coordinating Conjunctions	Prepositions	Conjunctive adverbs and transitions
Additive	N/A	and, or, nor, or else, either...or..., neither...nor..., not only... but (also)...	as well as, along with, together with, with	furthermore, besides, indeed, in addition, by the way, that is, alternatively, I mean, likewise, moreover, further, incidentally, what's more, in other words, similarly, for instance
Adversative	even though, although, though, while, whereas	but, yet, but...still, yet...still	despite, in spite of	however, instead, nonetheless, rather, nevertheless, in contrast on the other hand, in fact, on the contrary,
Causal	because, since, so that, as, inasmuch as, now that, such...that, so...that, in order that	so, for	because of, due to, in order to	therefore, hence, thus, consequently
Temporal	until, as soon as, after, as long as, once, before, whenever, when, while, since	and then	during, after, before, since, until, upon	then, next, after that, following that, before that, afterwards, meanwhile, beforehand
Condition	if, whether or not, unless, only if, providing (that), provided (that), in case	or (else)	N/A	otherwise

### B. Definition of "Use"

Since the present study intends to compare the use of EFL learners and native speakers, the definition of "use" is crucial for this study. Bolton et al. (2002) conducted a corpus-based study investigating students' use of logical connectors in academic writing. They defined "use" as "the identification of linguistic items as 'connectors', the measurement of the ratio of occurrence of connectors in our data, and the calculation of 'overuse' of connectors" (p. 173). Lee (2004) collected 202 academic essays written by Korean EFL learners totaling 214,363 words and analyzed their use of conjunctive adverbials, comparing them with several native corpora. He defined "use" as the frequency of the logical connectors, the types of functions in the discourse, and the syntactic positions in the sentences and drew some pedagogical implications from the overall similarities and differences of Korean EFL learners and native speakers for future academic writing teaching.

The research studies mentioned above reflected many aspects of academic writing needed improving, and the same problems may also exist in EFL learners' spoken discourse. In light of the research studies mentioned above, the present

study starts an exploration of the EFL learners' use of logical connectors, not in written discourse, but in spoken discourse. The "use" is thus examined from the following aspects: the raw frequency, the ratio of the frequency, the calculations of overuse, underuse, and the misuse of the logical connectors. The calculation method is provided in Section 3.4.

### *C. Corpus Linguistics and Its Use in English Language Teaching*

A corpus collects spoken or written texts or both, ranging from thousands of words to millions of words, which are stored and processed by computer and are readable in electronic forms (McCarthy & Carter, 2004). Corpus linguistics has been tightly associated with the research studies of many linguists, who carefully examined the concordance lines and collocations of phrases screen after screen, attempting to make sense of features of some languages from some seemingly orderly texts and collections (McCarthy & O'Keeffe, 2010). Because of the revolutionary and innovative features and use of corpora, corpus linguistics has become a linguistic methodology, which integrates piles of naturally occurring discourse into electronic sources. In other words, to investigate the features in spoken discourse, the research approaches within corpus linguistics call for collecting a substantial amount of naturally occurring data to uncover the frequency and occurrence of the grammatical and lexical items in different language settings and examines the patterns of their relationships (Hunston & Gill, 1998; Kennedy, 2014; McCarthy, 1991).

Although corpus linguistics is neither considered as a new field of linguistics nor a new theoretical foundation for English language studies, it has the potential to further develop other perspectives of language learning and teaching (Silva & Leki, 2004). Corpus linguistics provides an approach for empirical analysis of English language; by doing so, corpus linguistics would lead to the refinement and adjustment of our use of lexical and grammatical items. To be more specific, corpus linguistics is being applied in the pursuit of wider research fields such as English language teaching and learning, pragmatics, discourse analysis, sociolinguistics, literary stylistics, and health communications. By this token, the realms that corpus linguistics may lead to are far from lexis or grammar, which can be easily answered by computational linguistics, but as far and diverse as the research tools applied by second language acquisition and media studies (McCarthy & O'Keeffe, 2010).

Corpus linguistics is widely used in pedagogical applied linguistics. First, Tribble (2002) has been promoting Data-driven Learning (DDL) and applying the analyzed results to promote language teaching and learning. He has been bringing corpus data into the classroom and turning the traditional classroom to a student-oriented learning space where corpora provide knowledge bases, and students are controlling the pace of learning and questioning the teachers (McCarthy & O'Keeffe, 2010). Second, learner corpora, which are the collection of spoken and written learner language data, have shifted the focus from native speaker dominance to learner dominance or native speaker-learner contrastive studies. Third, another important field that corpus linguistics has an influence on is language testing and teacher training. Corpora can enhance professionalization by shedding light on scoring standards and rating rubrics. Corpora, with their featured wash-back implication, provide important empirical framework and benchmarks on which to base and grade examinations with a more international perspective (Barker, 2010; Farr, 2010). Fourthly, an increasing number of language teaching materials are becoming "corpus-informed" in a way that publishers of language teaching and learning such as Cambridge University Press, Oxford University Press, and Pearson-Longman are all sophisticatedly manipulating corpora and launching continuously updating teaching and learning materials (McCarthy & O'Keeffe, 2010, p. 8).

### *D. Learner Corpora*

As mentioned in the last section, the building of corpora has shifted its focus to the context of language learning, where learner language is one popular field. By collecting examples of learner use of language and comparing them with existing corpora, the features and drawbacks in learner language can be more thoroughly uncovered in a rather profound way than traditional work on error analysis and correction is (Bonelli, 2010). Computational learner corpora are defined as the collections of spoken or written data by EFL or ESL learner in a variety of language context in the electronic form (Granger, 2002). Research on learner corpora employs the principles and benchmarks from corpus linguistics and offers language description of learners, which will be later applied to language teaching pedagogy.

Among various learner corpora, learner spoken corpora with new and valuable data have also attracted linguists' renewed interest in the corpus linguistics studies in learners' spoken language. For example, Shirato and Stapleton (2007) compared Japanese learners' and native speakers' English spoken vocabulary and calculated and categorized some underused and overused logical connectors by Japanese learners. Furthermore, Wang (2008) investigated Chinese EFL learners' use of pragmatic markers and argued that Chinese learners tended to use logical connectors in the sentence-initial position.

This study presents a list of selective examples of ESL/EFL learner written and spoken corpora (see Table 2, for a more exhaustive list of learner corpora, see Granger (2002), Granger and Leech (2014), and (Tono, 2002)). The building of learner corpora makes it possible to uncover both qualitative and quantitative differences in linguistic and discourse features between EFL/ESL learners and native speakers.

TABLE II.  
A SELECTED LIST OF FREQUENTLY USED EFL LEARNER CORPORA

Corpus name	Abbr.	Attribution	Size (words)
Longman Learners' Corpus <sup>1</sup>	LLC	All level written essays	10,000,000
Cambridge Learners' Corpus <sup>2</sup>	CLC	All level written and spoken texts	10,000,000
International Corpus of Learner English <sup>3</sup>	ICLE	University-level written essays	3,700,000
Hong Kong University of Science and Technology Corpus of Learner English <sup>4</sup>	HKUST	University-level written essays	25,000,000
Chinese Learner English Corpus <sup>5</sup>	CLEC	All level written essays	1,000,000
Standard Speaking Test <sup>6</sup>	SST	All level spoken test samples	1,000,000
Louvain International Database of Spoken English Interlanguage <sup>7</sup>	LIDSEI	Higher intermediate to advanced learners' informal interviews	1,000,000
Michigan Corpus of Academic Spoken English <sup>8</sup>	MICASE	Learners' and native speakers' academic speech events	1,848,364
International Corpus Network of Asian Learners of English <sup>9</sup>	ICNALE	University-level written essays and spoken samples	1,800,000

### E. Methodological Approaches to Learner Corpora

One of the advantages of corpus-based analysis is its sufficiency in conducting both quantitative and qualitative analysis. For learner corpora analysis, the two frequently used methodological approaches are contrastive interlanguage analysis and computer-aided error analysis (Granger, 2002). The first method, as the method which has been employed in most of the Native Speakers (NS)/Non-native Speakers (NNS) and NNS/NNS corpus-based contrastive analysis, can accomplish both quantitative and qualitative comparisons. For another, the second method uses computer software to tag, retrieve, and analyze learners' errors and use of a language (Granger, 2002; Hunston & Gill, 1998).

The corpus-based contrastive analysis aims to:

[u]ncover the features of non-nativeness of learner language. At all levels of proficiency, but especially at the most advanced ones, these features will not only involve plain errors but differences in the frequency of use of certain words, phrases or structures, some being overused, others underused. (Granger & Leech, 2014, p. 13).

By this definition, corpus-based contrastive analysis pays more attention to L2 learners' use of particular lexical and grammatical items in comparison with those of native speakers. Multiple linguistic features such as discourse markers, complement clauses, adjective intensification, connectors, and direct questions have been investigated with the help of this approach. For example, Biber and Reppen (1998) investigated the coherence of marked themes in undergraduate students' essays by comparing the complement clauses in HKUST corpus of learner English with those in native English texts of *the Grammar of Spoken and Written English*. They found that Chinese learners prefer using connecting devices in the theme position while NS may place them more flexibly such as in sentence-final position.

Computer-aided error analysis closely examines learners' error in texts, using multiple major functions of computer programs. Considering that the conventional error analysis has been examining and criticizing the individual errors and giving feedback case by case, it gives negative feelings to the learner. By contrast, the computer-aided error analysis adopts a high degree of standardization, thus presenting the errors within the full context (Granger, 2002). The research can simply input an error-prone linguistic item such as a logical connector in this study and scan the corpus to search for all the cases of misuse with the help of the standardized text retrieval software such as *WordSmith* and *AntConc*.

### F. Research Questions

According to the research studies above, we can see that learners' overuse, underuse, and misuse of logical connectors in the written discourse have been closely examined by the comparison with the norms in the native speaker corpora, but few research studies were conducted on learner spoken discourse. This study intends to fill this research gap. Adopting the contrastive interlanguage analysis and computer-aided error analysis, this study first compiled a learner spoken corpus of TESOL students and then compared their use of logical connectors with those in MICASE. The present study examines the following research questions:

1. What are the overall frequencies of the logical connectors in the TESOL student corpora and MICASE?
2. What are the similarities and differences (i.e. underuse, overuse, and misuse) in using logical connectors between the TESOL students and native speakers?

## III. DATA COLLECTION AND RESEARCH METHODS

References (retrieved on June 2, 2017):

- <sup>1</sup> <http://www.pearsonlongman.com/dictionaries/corpus/learners.html>
- <sup>2</sup> <http://www.cambridge.org/gb/cambridgeenglish/better-learning/deeper-insights/linguistics-pedagogy/cambridge-english-corpus>
- <sup>3</sup> <https://uclouvain.be/en/research-institutes/ilc/cecl/icle.html>
- <sup>4</sup> [http://home.ust.hk/~autolang/download\\_WP.htm](http://home.ust.hk/~autolang/download_WP.htm)
- <sup>5</sup> <http://www.clal.org.cn/corpus/EngSearchEngine.aspx>
- <sup>6</sup> <http://www.alc.co.jp/edusys/refecorpus.htm>
- <sup>7</sup> <https://pul.uclouvain.be/book/?gcoi=29303100992060>
- <sup>8</sup> <https://quod.lib.umich.edu/cgi/c/corpus/corpus?page=home;c=micase;cc=micase>
- <sup>9</sup> <http://language.sakura.ne.jp/icnale/>

This section first explains how the participants were selected and the uniqueness of the triple identity of the participants. Then, the data collection procedure is introduced. Moreover, this section elaborates why this study decided to build a new spoken corpus instead of using the existing well-built corpora. Finally, it elaborates how the spoken data were transcribed into texts and analyzed by computer programs.

#### A. Participants

This study employs convenience sampling (Dörnyei, 2007), choosing the EFL TESOL students at a university in Hong Kong as the participants. The data collection event was conducted in research offices, study rooms, or any other convenient places in this university. The selected members of the target participants shared geographical proximity and easy accessibility in a certain period. All the participants were TESOL students in a university, and the time slots when they chose to participate in the spoken data collection event were between April 2017 to June 2017.

In addition to the ease and accessibility of sampling, according to Edge (1988), those participants share the same triple identity: language user, language teacher, and language analyst. First, as language users, those TESOL students have a high level of high English proficiency, which can be reflected in their language test results (see Chapter 4). Their adequate models in the classroom may inspire students to speak better-formed English and more accurate utterances because students also long for this language competence. Thus, students are more likely to use logical connectors more accurately if their teachers use them correctly. Second, according to the questionnaire survey result, 26% of them have full-time teaching experience, and 89% have part-time teaching experience. Significantly, 78% of them tend to pursue a teaching career after graduation, so the study on their use of logical connectors may contribute to their teaching performance as well as their capability of sensing their students' needs in the classroom. Thirdly, an individual report of their contrastive results with the native speakers were also sent to each participant. By analyzing their own performance in the spoken discourse, they are likely to rectify their inappropriate use and make better decisions on dealing with students' problems and difficulties on using logical connectors. As language analysts, the study provides a new perspective for those TESOL students that a comparison between students and native speakers can improve their teacher language awareness.

#### B. Spoken Data Collection

The data collection event took at most 25 minutes, including a retelling activity and filling out a questionnaire survey. The participants were first asked to watch a 5.5-minute video clip up to three times and simultaneously retell the plots in the last time when the video clip was playing. The video clip could be watched online on <https://youtu.be/vNJ5w-uQEEc>. The participants were not allowed to pause, reverse, or forward the entire or part of the video when the video clip was naturally playing. They were not allowed to write down or mark anything on paper or other devices. Those steps were to guarantee that the spoken discourse collected were all spontaneously and naturally occurring data. The spoken discourse of the participants was audio-recorded by the researcher's mobile phone, which was safely encrypted. After retelling, the participants were asked to complete a questionnaire investigating their demographical information (see Appendix 1).

The video clip was excerpted from one of the Chaplin series *The Circus*. There was no dubbing or subtitle in the video clip; therefore, the participants could not find any hint or tip on the English language use. The video clip acted only as a reminder of how the storyline was so that the participants could continuously deliver the naturally occurring spoken discourse without long intervals or intermittent speechlessness.

#### C. The Corpora

There are two corpora used for the research, namely MICASE and the newly built EFL TESOL student spoken corpus. The English native speaker corpus is the existing corpus—MICASE, which contains 1,848,364 words transcribed from 152 pieces of about 200 hours of spoken discourse from 1,571 speakers aged from 17 to 30. MICASE focuses on the contemporary university-level English speech at the University of Michigan. The places of the speech recording were all on the campus such as classrooms, discussion rooms, meeting rooms, and labs.

The EFL TESOL student spoken corpus is constituted by the spoken samples from 27 samples of 12,241 words in total from 27 Chinese students, including students from Mainland China and HKSAR. After obtaining the spoken samples from the participants, the samples were verbatim transcribed into separated files named by their order of participation. As the counterpart of MICASE, the current corpus uses the same way of transcribing the spoken data, each file containing time, location, participant, duration, language background, and the texts. Those data are used to retrieve tokens and frequencies of logical connectors and to compare with the filtered NS corpus of MICASE.

There are several reasons why a new spoken corpus is required. Firstly, even though CLC, SST, and ICNALE are all great spoken corpora with verbatim transcriptions of a substantial number of spoken samples from EFL learners from all over the world, some of them are not free of charge, and others are not easily accessible. Furthermore, among the accessible corpora, the percentage of the participants who speak advanced English in the corpora mentioned above is relatively low. For example, among 950 EFL/ESL participants who gave their spoken samples to ICNALE, only 160 (16.84%) of them are above B2 according to Common European Framework of Reference (CEFR) (Council of Europe, 2001), whose spoken discourse is not much of meaning for this study. The data needed for this study should be advanced English learners whose English proficiency ranges among C1 or above. According to the demographical

information from the questionnaire, 100% of those TESOL students have obtained C1 or higher level. Thirdly, the EFL TESOL students have the triple identity, making them distinctive and unique for this study, as is discussed in the last sections.

#### D. Analysis

The present study focuses on the analysis of logical connector use in the learners' and native speakers' spoken discourse. Specifically, the ratio of the frequency and the overuse and underuse are measured by quantitative methodology. In addition, the misuse is presented by raising typical examples of how the TESOL students misused some certain logical connectors.

For the TESOL student corpus, the list of logical connectors did not simply duplicate the existing classifications. Instead, during the process of verbatim transcription, all the logical connectors were manually marked up, and all the appearance of the logical connectors in the transcriptions was checked twice to ensure the reliability and accuracy of the research. This method is advantageous in that it provides a consistent and non-arbitrary list of logical connectors. The application of this "accurate and contextualized" list of logical connectors has greatly improved the efficacy of the following examination, for it makes practicable to employ the same list as the standard in the calculation of "overuse" and "underuse".

For MICASE, the list of logical connectors classified by Sorensen (1997) and Halliday (1976, pp. 242-243) was adopted. Then, the two lists were combined into the full list in Table 3. Unavoidably, some words have more than one speech. For example, *for* can be a preposition (e.g. *She brewed some coffee for me.*) and a coordinating conjunction (e.g. *You'd better take this medicine, for you are ill.*). This study adopted the results retrieved from the online program CLAWS part-of-speech post-tagger for English to distinguish the part of speech.

As the inventory of the logical connectors in English is very large, the free-of-charge computer program AntConc (3.4.4m) helped in locating concordance, collocations, terms, and key characters in the transcribed texts. This quantitative methodology first counted how many logical connectors were employed in both corpora and then divided these numbers by the total words of each corpus.

For example, *however* in the TESOL student corpus appeared 23 times; then, the Frequency of *however* was 23, the Ratio of Frequency  $23/12,241=0.0018791$  (1.88‰), RF per 1000 words 1.88. Adopting the same calculation method, we know that the RP per 1000 words of *however* in MICASE was 0.12. Then, we can get the difference between 1.88 and 0.12 was 1.76, which was a positive value denoting the overuse of *however* in the TESOL student corpus. The overuse and the underuse were fundamentally based on the comparison of two corpora by this calculation method.

### IV. FINDINGS AND DISCUSSION

#### A. Overall Frequencies of Logical Connectors

Table 3 demonstrates a full list of the 98 logical connectors found in both corpora. There were 52 logical connectors found in the TESOL student corpus and 72 in MICASE. From the table, we can see that the native speakers used a much wider range of logical connectors than the TESOL students. Specifically, the native speakers tended to use more types of logical connectors such as *so*, *therefore*, *so that*, and *since* to express causal relationship in the sentence-level spoken discourse while the TESOL students mainly used *so* and *because* to express the causal relationship.

For the TESOL students, the low frequencies in the types of the logical connectors and the high frequencies of the same tokens indicate the same types of the logical connectors were highly repetitive. As Milton (1998) suggested, the EFL learners tend to use an exclusion of expressions which they consider that they know and can manage. This may be also true of the TESOL students' use of logical connectors. They used a smaller repertoire of logical connectors than the native speakers did, but in order to keep the coherence of the spoken discourse, they thus increased the frequency of the same logical connector.

TABLE III.  
ALL LOGICAL CONNECTORS IN THE TESOL STUDENTS SPOKEN CORPUS AND MICASE (FREQ. = FREQUENCY, RF = RATIO OF FREQUENCY, A POSITIVE VALUE DENOTES OVERUSE, A NEGATIVE VALUE DENOTES UNDERUSE)

Logical Connectors	TESOL Corpus		(+/-)	MICASE	
	Freq.	RF per 1000 words		Freq.	RF per 1000 words
<b>Coordinating Conjunctions</b>					
<i>and</i> (excluding "and then")	595	48.61	23.19	46976	25.41
<i>but</i>	196	16.01	9.55	11952	6.47
<i>so</i> (excluding "so that")	156	12.74	2.68	18598	10.06
<i>or</i>	8	0.65	-4.06	8707	4.71
<i>yet</i>	4	0.33	0.09	438	0.24
<i>for</i>	0	0.00	-0.05	92	0.05
<i>nor</i>	0	0.00	-0.02	37	0.02
<b>Conjunctive Adverbs and Transitions</b>					
<i>and then</i>	60	4.90	3.15	3242	1.75
<i>then</i> (excluding "and then")	36	2.94	-0.47	6301	3.41
<i>suddenly</i>	27	2.21	2.17	66	0.04
<i>however</i>	23	1.88	1.76	214	0.12

Logical Connectors	TESOL Corpus		(+/-)	MICASE	
	Freq.	RF per 1000 words		Freq.	RF per 1000 words
<i>unfortunately</i>	17	1.39	1.36	61	0.03
<i>accidentally</i>	16	1.31	1.30	6	0.00
<i>finally</i>	13	1.06	0.97	175	0.09
<i>at this time</i>	12	0.98	0.96	36	0.02
<i>at that time</i>	7	0.57	0.55	49	0.03
<i>immediately</i>	7	0.57	0.53	75	0.04
<i>luckily</i>	5	0.41	0.41	5	0.00
<i>fortunately</i>	4	0.33	0.32	14	0.01
<i>at the moment</i>	4	0.33	0.31	28	0.02
<i>unluckily</i>	3	0.25	0.24	1	0.00
<i>thankfully</i>	3	0.25	0.24	1	0.00
<i>at the same time</i>	2	0.16	0.07	165	0.09
<i>at first</i>	2	0.16	0.14	52	0.03
<i>by the way</i>	1	0.08	0.02	118	0.06
<i>incidentally</i>	1	0.08	0.08	8	0.00
<i>absolutely</i>	1	0.08	0.00	149	0.08
<i>afterwards</i>	1	0.08	0.06	49	0.03
<i>basically</i>	1	0.08	-0.43	948	0.51
<i>meanwhile</i>	1	0.08	0.08	6	0.00
<i>surprisingly</i>	1	0.08	0.07	14	0.01
<i>unexpectedly</i>	1	0.08	0.08	0	0.00
<i>at this moment</i>	1	0.08	0.08	5	0.00
<i>therefore</i>	1	0.08	-0.03	201	0.11
<i>instead (excluding "instead of")</i>	1	0.08	0.04	82	0.04
<i>instantly</i>	1	0.08	0.08	6	0.00
<i>furthermore</i>	0	0.00	-0.01	15	0.01
<i>in addition (excluding "in addition to")</i>	0	0.00	-0.01	16	0.01
<i>moreover</i>	0	0.00	-0.01	18	0.01
<i>what's more</i>	0	0.00	0.00	4	0.00
<i>for instance</i>	0	0.00	-0.08	142	0.08
<i>"I mean,"</i>	0	0.00	-0.25	455	0.25
<i>in other words</i>	0	0.00	-0.12	229	0.12
<i>alternatively</i>	0	0.00	-0.01	11	0.01
<i>indeed</i>	0	0.00	-0.07	122	0.07
<i>similarly</i>	0	0.00	-0.01	15	0.01
<i>likewise</i>	0	0.00	0.00	7	0.00
<i>"that is,"</i>	0	0.00	-0.06	120	0.06
<i>besides</i>	0	0.00	-0.03	59	0.03
<i>for example</i>	0	0.00	-0.21	381	0.21
<b>Prepositions</b>					
<i>with</i>	20	1.63	-2.60	7829	4.24
<i>instead of</i>	3	0.25	0.11	251	0.14
<i>actually</i>	3	0.25	-1.09	2469	1.34
<i>despite</i>	1	0.08	0.07	26	0.01
<i>because of</i>	1	0.08	-0.11	352	0.19
<i>without</i>	1	0.08	-0.12	366	0.20
<i>due to</i>	0	0.00	-0.06	107	0.06
<i>in fact</i>	0	0.00	-0.26	488	0.26
<i>in addition to</i>	0	0.00	-0.02	38	0.02
<i>during</i>	0	0.00	-0.16	297	0.16
<i>in spite of</i>	0	0.00	0.00	8	0.00
<i>upon</i>	0	0.00	-0.05	91	0.05
<i>concerning</i>	0	0.00	-0.01	15	0.01
<i>regarding</i>	0	0.00	-0.02	29	0.02
<i>within</i>	0	0.00	-0.21	391	0.21
<b>Subordinating Conjunctions</b>					
<i>because (excluding "because of")</i>	20	1.63	-0.65	4227	2.29
<i>after (prep.+conj.)</i>	14	1.14	0.64	940	0.51
<i>when</i>	10	0.82	-1.39	4083	2.21
<i>in order (not) to</i>	8	0.65	0.50	289	0.16
<i>if</i>	8	0.65	-4.56	9644	5.22
<i>as</i>	5	0.41	-0.84	2302	1.25
<i>while (temporal+adversative)</i>	5	0.41	0.16	457	0.25
<i>before (prep.+conj.)</i>	4	0.33	-0.29	1135	0.61
<i>in case (that)</i>	2	0.16	0.14	35	0.02
<i>until</i>	1	0.08	-0.11	350	0.19
<i>so that</i>	1	0.08	-0.78	1584	0.86
<i>since (prep.+conj.)</i>	0	0.00	-0.27	505	0.27
<i>until (prep.+conj.)</i>	0	0.00	-0.19	350	0.19
<i>as soon as</i>	0	0.00	-0.03	53	0.03
<i>as long as</i>	0	0.00	-0.05	91	0.05
<i>as far as</i>	0	0.00	-0.08	143	0.08

Logical Connectors	TESOL Corpus		(+/-)	MICASE	
	Freq.	RF per 1000 words		Freq.	RF per 1000 words
<i>though</i>	0	0.00	-0.44	814	0.44
<i>although</i>	0	0.00	-0.13	240	0.13
<i>whereas</i>	0	0.00	-0.09	158	0.09
<i>now that</i>	0	0.00	-0.08	139	0.08
<i>by the time</i>	0	0.00	-0.02	36	0.02
<i>unless</i>	0	0.00	-0.11	199	0.11
<i>anyway</i>	0	0.00	-0.17	320	0.17
<i>whether</i>	0	0.00	-0.38	694	0.38
<b>TOTAL</b>	<b>1327</b>			<b>139622</b>	

### B. Overuse and Underuse of Logical Connectors

The examination of the individual logical connectors indicates the underuse and overuse of logical connectors. Table 4 and Table 5 demonstrate the ten most overused and the ten underused logical connectors in both corpora.

TABLE IV.  
THE TOP 10 MOST OVERUSED LOGICAL CONNECTORS

Logical Connectors	TESOL Corpus		(+/-)	MICASE	
	Freq.	RF per 1000 words		Freq.	RF per 1000 words
<i>and (excluding "and then")</i>	595	48.61	23.19	46976	25.41
<i>but</i>	196	16.01	9.55	11952	6.47
<i>and then</i>	60	4.90	3.15	3242	1.75
<i>so (excluding "so that")</i>	156	12.74	2.68	18598	10.06
<i>suddenly</i>	27	2.21	2.17	66	0.04
<i>however</i>	23	1.88	1.76	214	0.12
<i>unfortunately</i>	17	1.39	1.36	61	0.03
<i>accidentally</i>	16	1.31	1.30	6	0.00
<i>finally</i>	13	1.06	0.97	175	0.09
<i>at this time</i>	12	0.98	0.96	36	0.02
<b>TOTAL</b>	<b>1150</b>			<b>99466</b>	

TABLE V.  
THE TOP 10 MOST UNDERUSED LOGICAL CONNECTORS

Logical Connectors	TESOL Corpus		(+/-)	MICASE	
	Freq.	RF per 1000 words		Freq.	RF per 1000 words
<i>if</i>	8	0.65	-4.56	9644	5.22
<i>or</i>	8	0.65	-4.06	8707	4.71
<i>with</i>	20	1.63	-2.60	7829	4.24
<i>when</i>	10	0.82	-1.39	4083	2.21
<i>actually</i>	3	0.25	-1.09	2469	1.34
<i>as</i>	5	0.41	-0.84	2302	1.25
<i>so that</i>	1	0.08	-0.78	1584	0.86
<i>because (excluding "because of")</i>	20	1.63	-0.65	4227	2.29
<i>then (excluding "and then")</i>	36	2.94	-0.47	6301	3.41
<i>though</i>	0	0.00	-0.44	814	0.44

From Table 4, the TESOL students' ten most frequently used logical connectors comprise 87% (1150/1327) of the total number of the logical connectors; by contrast, these ten logical connectors only account for 71% (99466/139622) of the total number of the native speakers' use of logical connectors. The repeated tokens of the same kind of logical connectors have inflated the frequencies of the logical connectors with an overwhelming percentage of 87%, which one more time indicates that the TESOL students tended to more frequently use the same types of logical connectors than the native speakers did. Then, we can notice that some logical connectors such as *suddenly*, *at this time*, and *accidentally* which were very frequently used by the TESOL students were rarely observed in MICASE. Furthermore, from Table 4, we can see that in the TESOL student corpus the logical connectors *and*, *but*, *and then*, and *so* were particularly overused. Those four logical connectors, comprising 76% (1007/1327) of all, are all coordinating conjunctions, while those four logical connectors only account for 58% (80768/139622) of the native speakers' logical connector use. It seems that the TESOL students gave coordinating conjunctions higher priority while doing the retelling. Examples of overuse of an additive coordinating conjunction *and* is illustrated in the following excerpt from the TESOL student corpus. Grammatical and lexical errors are in the originals.



**Excerpt 1:**

(TESOL-Corpus-016: The transcribed text from 3'52" to 4'22" by a student at a university in Hong Kong, recorded in a study room, Mandarin as the first language, English as the foreign language)

**And** the woman did that, **and** the woman say, "hurry you can go out now, hurry up." **And** the man just pretend that she uh he was not afraid of that lion **because** that lion he thought the lion was quite timid and not very fierce, **so** he just want to tease the lion, **and** the lion just roars at him, **and** she just suddenly ran out of the cage, **and** the woman locked the door, **and** the woman also ran away, **and** she saw the man was climbing the man climb onto a maybe a stick or something, **and** on the top of it, **and** the woman said: "you you you're not afraid now?"

From the excerpt, we notice that this TESOL student initiated almost every independent clause with an additive coordinating conjunction *and*. Leech (2000, p. 676) concluded this phenomenon as the characteristics of spoken grammar that the learners in spoken discourse may have a constrained lexico-grammatical repertoire and may repetitively use the items in it. Even though it is not grammatically incorrect that independent clauses can be added by additive coordinating conjunctions endlessly, the hearer may find it difficult to disentangle the inner logic of the speech when the topic is switching. However, for the case of overuse of *and* in spoken discourse, Willis (2003, p. 186) held that learners use *and* quite a lot to give a stream of information in a more linear way, making the listener easier to follow within a focused topic.

Table 5 demonstrates the TESOL students' underuse of ten logical connectors, most of which were subordinating conjunctions used in adverbial clauses. First, the native speakers more frequently used *if* in spoken language. *If* can be used in conditional sentences or starts an object clause. In MACASE, both of two uses appeared, but in the TESOL student corpus, all *if* started object clauses (See Fig. 1). Surprisingly, after the word-for-word examination, there was also no other logical connectors used for conditional meaning in the TESOL student corpus. This is probably because the participants had little time to produce more complicated conditional sentences in the dynamic and transient spoken discourse after receiving the abundant streaming information from the video clip. Second, the underuse of the causal subordinating conjunction *so that* confirmed the previous deduction that the TESOL students gave the logical connectors except the coordinating conjunctions such as *so* a very low priority to convey causal relationship. Third, *actually* and *basically* were also greatly underused by the TESOL students. *Actually* is "used to stress that a statement is true especially when it differs in some way from what might have been thought or expected" (Merriam-Webster, N.D.-a), and *basically* is "used to show that a statement is expressing the most important reason for something" (Merriam-Webster, N.D.-b), both of which are used to stress the previous statement to achieve a certain aim. Subjectively speaking, the TESOL students probably did not have enough time to proceed a repetition or a rephrasing after delivering part of speech when receiving the abundant streaming information in the video clip, or they did not have the capability to reproduce information in the shortest possible time. Objectively speaking, the genre of the retelling task might have discouraged their needs to speak such clauses.

... he takes out his handkerchief and wants to see **if** there is someone come comes to his ...  
 ... so he wakes his neck looking to see **if** there is anyone to help, but nobody ...  
 ... it was it won't make the lions wake **if** he just cover his own ears, but the dog ...  
 ... And then he whips the his scarf to find **if** there is someone that could offer him help ...  
 ... He took out his napkin to check out **if** there was somebody to help him ...  
 ... the man immediately closed the door and check **if** the lion was uh wake up.  
 ... a handkerchief from his pocket and waved to see **if** anyone is there, but there is no one ...  
 ... and the man took out a scarf to see **if** there is anyone who can help him ...

Figure 1. All concordance lines with the logical connector *if* in the TESOL student corpus. (Concordance lines are process by AntConc 3.4.4m)

### C. Misuse of Logical Connectors

The logical connector *in case*, which appeared only twice in the TESOL student corpus, was misused according to its definition in Webster Online Dictionary. *In case* should be used "as a precaution against an event" (Merriam-Webster, N.D.-c); the collocations of *in case* are normally "in case of + noun" and "in case (that) + a clause", but one of the participants used it in "... and get out of the room quietly **in case to\*** wake the lion up..." The appropriate use should be "in case that the lion would wake up". The reason why the participant misused *in case* may be that the use of *in case* and *in order not to* were mixed up due to L1 transfer. *In order not to* means "為了不" in Chinese while *in case* means "以防" (to prevent something from happening). The latter contains a subjunctive mood, but Chinese does not distinguish them a lot in spoken discourse, making the participant believe that *in case to\** was equal to *in order not to*.

There are three findings from the results above. First, the EFL TESOL students tended to use a smaller set of logical connectors but use them more frequently than the English native speakers. Second, additive coordinating conjunctions such as *and*, *so*, and *but* were the most overused logical connectors by the TESOL students. Their preference of employing a smaller repertoire of logical connectors seemed to lead to the highly repetitive use of the same logical connectors in their spoken discourse, which might make their speech redundant, dull, and less interesting. Third, the most underused logical connectors such as *if*, *when*, *so that*, and *though* show that the TESOL students did not tend to or were not encouraged by the retelling task to speak adverbial clauses and further process the inner connections in the discourse-level.

The first finding echoes that in Fung and Carter's (2007) study, which investigated the range and frequencies in EFL learners' discourse marker use in spoken English and found that learners used a restricted range of DMs with a high frequency of particular markers. They deemed this unnatural use of markers resulted from the unnatural linguistic input and the traditional grammar-translation teaching method. Many locally-produced textbooks were noticed to attach emphasis only on the details of logical connector use and prepositional content rather than on the pragmatic use. For the second finding, English learners of Hong Kong Cantonese speakers (Fung & Carter, 2007) and Chinese speakers (Wang, 2008) were all found to overuse coordinating conjunctions (e.g. *and*, *but*, and *so*) and conjunctive adverbs and transitions (e.g. *however*, and *finally*) in spoken discourse. For the third finding, Schleppegrell (1996) did find that the frequencies of the adverbial clauses in ESL learners' spoken English were fewer than those in written discourse, but there was no explicit finding illustrating that EFL/ESL learners use fewer temporal adverbials and adverbial clauses than native speakers, which needs further investigation.

## V. CONCLUSION

Corpus linguistics, especially the corpus-based research methodology, would greatly contribute to the understanding of second language acquisition and the improvement of English-speaking teaching. It provides us with insights into EFL learners' spoken discourse, thus deepening our understanding of learner's language and how each item is actually used for communication. The present study demonstrates how the comparison between the learner corpus and the native speaker corpus can be employed to investigate learners' difficulties and develop corresponding teaching strategies. However, much more research needs doing until more appropriate teaching strategies and materials are established. In addition, corpus-based interlanguage contrastive studies into logical connectors would, in turn, contribute to the compilation of better learner dictionaries with more accurate and contextualized use (Granger & Tyson, 1996).

Some pedagogical implications can also be drawn from this study now that there were noticeable discrepancies between the EFL TESOL students and the native speakers. First, for teachers, it is crucial to attach more emphasis on contextualization of logical connectors in authentic texts. As Zamel (1984) mentioned, learners have to learn the detailed differences among individual semantic linking devices even though getting across each logical connector in sentence-level, passage-level, and discourse-level is challenging and time-consuming. Second, I agree with Crewe (1990) that the inter-switchable lists of logical connectors in many textbooks are quite misleading and thus are necessary to be rectified with all our might. I also agree with Zamel (1984, p. 116) that "learning when not to use them (connectors) is as important as learning when to do so". The group of the TESOL students need to in particular be cautious of the inappropriateness of overusing additive logical connectors. This requires them to search and synthesize more relevant references on the meaning, forms, and uses of logical connectors. Third, as Wierzbicka (2003) stated, logical connectors, which are important elements demonstrating one's cultural, social, and linguistic capability, are not just aimed at providing discourse cohesiveness and communication efficiency, but also at enhancing interactive and multi-cultural collaboration. This paper advocates incorporating logical connectors into the language curriculum not only as them per se but as a tool to improve learners' pragmatic competence in English speaking. We need to help learners acquire fluent and naturalistic conversational skills and a sense of security when speaking L2.

One limitation of the study emerged from the relatively small size of the TESOL student corpus, compared with that of MICASE. For the underexplored area of the logical connector use in the spoken discourse, the larger the TESOL student spoken corpus is, the stronger the conclusions we may draw from the contrastive analysis. Furthermore, the tokens of the logical connectors drawn from MICASE should not be regarded as an exhaustive list. Now that the study of logical connectors in spoken discourse is still in its infancy, an exhaustive and more consentaneous taxonomy with an actual number to be calculated in a legitimate category still needs to be developed. Another limitation may result from the genre of the retelling task. The nature of the task might not have allowed or encouraged the participants to use some certain logical connectors (e.g. *if* in conditional sentences and *so that* to express causal relationship); by simply retelling a story, they did not need to make any hypothetical or rephrased statements in adverbial clauses. My hypothesis is that the genre of the retelling task to some extent influences the logical connectors to be used; for example, additive adverbials may be more frequently used when expounding argument and supporting ideas in argumentation while temporal adverbials may be more employed in storytelling like the assigned task in this research.

Ghasemi (2013) confirmed the relevance between writing genres and logical connector use. Accordingly, further exploration can be focused on the relationship between the types of the logical connectors used and the genre of the speaking samples. From Table 1 and Table 3, it can be noticed that the temporal logical connectors such as *when*, *while*, and *suddenly* were much more frequently used by the TESOL students than the other types, but this preliminary finding cannot be the sufficient condition to prove that the genre of the task influences the use of logical connectors in spoken discourse. Further researchers can assign retelling tasks of different genres such as storytelling, argumentation, and exposition to the participants and check whether there is a correlation between genre and logical connector use.

## APPENDIX

THE UNIVERSITY OF HONG KONG  
Faculty of Education and Center for Applied English Studies

No. \_\_\_\_\_

THE CONTENTS OF THIS FORM ARE CONFIDENTIAL. INFORMATION IDENTIFYING THE RESPONDENT WILL NOT BE DISCLOSED UNDER ANY CIRCUMSTANCES.

**Introduction**

This survey is conducted by Wu Hao at MA(TESOL) program at the University of Hong Kong to better understand the features in your English spoken discourse. It should take you about three minutes to complete this questionnaire. Please give your answers sincerely as only this will guarantee the success of the investigation. This questionnaire is aimed at investigating your demographic information.

If you have any question, please feel free to contact Mr. WU Hao, Chuck by email ([wuhaodlc@connect.hku.hk](mailto:wuhaodlc@connect.hku.hk)). Thank you very much for your help. For further information of confidentiality, data retention, and participation and withdrawal, please refer to the Informed Consent Form.

**Part 1: Demographic Information**

This section is concerned with your backgrounds information.

- (a) **Name:** \_\_\_\_\_
- (b) **Gender:**  Male  Female
- (c) **Age:** \_\_\_\_\_
- (d) **Nationality:** \_\_\_\_\_
- (e) **First Language:**  Cantonese  Mandarin  other Chinese dialects (please specify \_\_\_\_\_)  
 English  Others (please specify \_\_\_\_\_)
- (f) **English proficiency:**  
IELTS: Overall: \_\_\_\_\_; Speaking: \_\_\_\_\_  
TOEFL: Overall: \_\_\_\_\_; Speaking: \_\_\_\_\_  
Others (please specify \_\_\_\_\_)
- (g) English Teaching Experience: Full-time: \_\_\_\_\_; Part-time: \_\_\_\_\_
- (h) Would you like to be an English teacher after graduation?  Yes  No  Not decided

\*\*\*\*\* This is the end of the questionnaire. Thank you very much. \*\*\*\*\*

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Dr. YUNG Wai Ho Kevin at The Education University of Hong Kong for always offering patient instructions, valuable comments, and kind encouragement while I was completing this project. Also, I am grateful to the journal editors for all the insightful comments, which have much improved the paper.

## REFERENCES

- [1] Altenberg, B., & Tapper, M. (1998). The use of adverbial connectors in advanced Swedish learners' written English. In S. Granger (Ed.), *Learner English on computer* (pp. 80-93). London: Longman.
- [2] Barker, F. (2010). How can corpora be used in language testing? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics*. New York: Routledge.
- [3] Biber, D., & Reppen, R. (1998). Comparing native and learner perspectives on English grammar: A study of complement clauses. In S. Granger (Ed.), *Learner English on computer* (pp. 145-158). London: Longman.
- [4] Bolton, K., Nelson, G., & Hung, J. (2002). 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. doi:10.1075/ijcl.7.2.02bol
- [5] Bonelli, E. T. (2010). Theoretical overview of the evolution of corpus linguistics. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics*. New York: Routledge.
- [6] Chafe, W. L. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
- [7] Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- [8] Crewe, W. J. (1990). The illogic of logical connectives. *ELT Journal*, 44(4), 316-325. doi:10.1093/elt/44.4.316.
- [9] Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.
- [10] Edge, J. (1988). Applying linguistics in English language teacher training for speakers of other languages. *ELT Journal*, 42(1), 9-13.
- [11] Farr, F. (2010). How can corpora be used in teacher education? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge*

*handbook of corpus linguistics*. New York: Routledge.

- [12] Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics*, 28(3), 410-439. doi:10.1093/applin/amm030.
- [13] Ghasemi, M. (2013). An investigation into the use of cohesive devices in second language writings. *Theory and Practice in Language Studies*, 3(9), 1615. doi:10.4304/tpls.3.9.1615-1623.
- [14] Granger, S. (2002). A bird's eye view of computer learner corpora. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). Philadelphia: John Benjamins Pub.
- [15] Granger, S., & Leech, G. (2014). *Learner English on computer*. New York: Routledge.
- [16] Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non - native EFL speakers of English. *World Englishes*, 15(1), 17-27.
- [17] Green, C. F., Christopher, E. R., & Mei, J. L. K. (2000). The incidence and effects on coherence of marked themes in interlanguage texts: A corpus-based enquiry. *English for Specific Purposes*, 19(2), 99-113.
- [18] Halliday, M. A. K. (1976). *Cohesion in English*. London: Longman.
- [19] Hunston, S., & Gill, F. (1998). Verbs observed: A corpus-driven pedagogic grammar. *Applied Linguistics*, 19(1), 45-72.
- [20] Kennedy, G. (2014). *An introduction to corpus linguistics*. New York: Routledge.
- [21] Kusuyama, Y. (2003). *How Japanese ESL writers begin sentences in academic writing*. (Ph.D.), University of California, Los Angeles, Los Angeles.
- [22] Lee, E.-J. (2004). A corpus-based analysis of the Korean EFL learners' use of conjunctive adverbials. *English Teaching*, 59(4), 283-301.
- [23] Leech, G. (2000). Grammars of spoken English: New outcomes of corpus - oriented research. *Language Learning*, 50, 675-724. doi:10.1111/0023-8333.00143.
- [24] McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge: Cambridge University Press.
- [25] McCarthy, M., & Carter, R. (2004). Introduction. *Journal of Pragmatics*(36), 147-148.
- [26] McCarthy, M., & O'Keeffe, A. (2010). Historical perspective: What are corpora and how have they evolved? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics*. New York: Routledge.
- [27] Merriam-Webster. (N.D.-a). Actually. Retrieved from [www.merriam-webster.com/dictionary/actually](http://www.merriam-webster.com/dictionary/actually) (assessed Feb 24 2019).
- [28] Merriam-Webster. (N.D.-b). Basically. Retrieved from [www.merriam-webster.com/dictionary/basically](http://www.merriam-webster.com/dictionary/basically) (assessed Feb 24 2019).
- [29] Merriam-Webster. (N.D.-c). In case. Retrieved from [www.merriam-webster.com/dictionary/in%20case](http://www.merriam-webster.com/dictionary/in%20case) (assessed Feb 24 2019).
- [30] Milton, J. (1998). Exploiting L1 and interlanguage corpora in the design of an electronic language learning and production environment. In S. Granger (Ed.), *Learner English on computer* (pp. 186-198). London: Longman.
- [31] Schleppegrell, M. J. (1996). Conjunction in spoken English and ESL writing. *Applied Linguistics*, 17(3), 271-285. doi:10.1093/applin/17.3.271.
- [32] Shirato, J., & Stapleton, P. (2007). Comparing English vocabulary in a spoken learner corpus with a native speaker corpus: Pedagogical implications arising from an empirical study in Japan. *Language Teaching Research*, 11(4), 393-412. doi:10.1177/1362168807080960.
- [33] Silva, T., & Leki, I. (2004). Family matters: The influence of applied linguistics and composition studies on second language writing studies—Past, present, and future. *The Modern Language Journal*, 88(1), 1-13.
- [34] Sorensen, M. N. (1997). Logical Connectors. Retrieved from <https://staff.washington.edu/marnell/grammar/logicalconnectors.html> (assessed Feb 24 2019).
- [35] Tono, Y. (2002). *The role of learner corpora in SLA research and foreign language teaching: the multiple comparison approach*. (Ph.D.), University of Lancaster, Lancaster.
- [36] Tribble, C. (2002). Corpora and corpus analysis: New windows on academic writing. In J. Flowerdew (Ed.), *Academic discourse* (pp. 131-149). London: Longman.
- [37] Wang, L. (2008). *An investigation into pragmatic markers in Chinese EFL learners in spoken English: A corpus-based approach*. (Ph.D.), Shanghai Jiaotong University, Shanghai.
- [38] Wierzbicka, A. (2003). *Cross-cultural pragmatics: The semantics of human interaction*. Berlin: Walter de Gruyter.
- [39] Willis, D. (2003). *Rules, patterns and words: Grammar and lexis in English language teaching*. Cambridge: Cambridge University Press.
- [40] Zamel, V. (1984). Teaching those missing links in writing. In S. McKay (Ed.), *Composing in a Second Language* (pp. 110-122). San Francisco State University, Cambridge: Newbury House.



**Hao Wu**, born in Dalian, Liaoning Province, China in 1993. He graduated from Faculty of Education of The University of Hong Kong and obtained the degree of Master of Arts in Teaching English to Speakers of Other Languages.