

A Corpus-Based Study between L1 and L2 English Academic Experts: Linking-Adverbial Usage in Linguistics Research Articles

Chunxiao Yue

English Department, City University of Hong Kong, Hong Kong, China

Email: chunxyue@um.cityu.edu.hk

Manuscript received January 8, 2024; revised February 15, 2024; accepted March 9, 2024; published May 24, 2024

Abstract—This corpus-based comparative study probes into how Linking Adverbials (LAs hereafter) are employed disparately in Chinese and native professional writers' research articles of the linguistics discipline. To achieve this, two corpora of research articles composed by Chinese and English native experts are manually compiled, followed by a frequency-plus-difference test as the quantitative analysis and a subjective interpretation as the qualitative analysis. Several findings are revealed in this study: in view of using frequency, Chinese linguistics professionals demonstrate differences in over half of the interested LAs (32 in sum) when compared to English natives. In view of using appropriateness, it is unveiled that not only do Chinese experts overtly use most of the additive and causal/resultative LAs, but they underuse all discussed adversative LAs. There are also some fresh spots where misuse runs through the whole analysis and takes place in all LAs categories, and some LAs usage problems are common in writers with diversified biographic backgrounds including but not limited to non-English L1, academic stage, and linguistic proficiency. With these discussed, the present study culminates in providing new insights into the corpus-assisted instruction and training of English research paper writing for professionals.

Keywords—linking adverbials, research articles, academic English writing, cohesion

I. INTRODUCTION

In the scope of English academic writing, there is a vast array of corpus-based studies on lexico-grammatical features [1–5]. Among a wide range of grammatical devices, Linking Adverbials (LAs) feature with the function of promoting the overall academic writing quality with the enhancement of text coherence, cohesion, organization, and logical-relationships expression, which has been proved in plenty of substantial studies as early as decades ago [1, 6–8]. In recent years, more concrete comparative research on LAs usage has been thriving, some of which are cross-register [9, 10] or cross-discipline works [11, 12], while others concern its use in varied student colonies [13, 14]. However, scarcely does comparative research directly inquire into LAs usage by Chinese versus native experts in linguistics discipline, in which higher criteria of academic writing are entailed. Also, in spite of the fact that Chinese linguistic experts are capable of exerting grammatical devices in a comparatively sophisticated way, there must be some room for them to make headways in perfecting the use of LAs and enhancing the holistic coherence and cohesion of paper, thereby strengthening the overall quality in research paper writing. In this scenario, it is a must to do research that helps elicit the overuse, underuse, and misuse of LAs by Chinese professionals in their research papers. Taking a different

view, the current study strives to investigate the academic writing quality of Chinese linguistics professionals by investigating the distinctions of their ways of deploying linking adverbials compared with native writers. A contrastive study is thus conducted on the basis of two representative corpora (Chinese professionals corpus and native professionals corpus) including 1.03 million tokens, with a hope that robust statistical results and well-rounded manual interpretation can join hands to dig out some earlier-concealed findings, shedding some new lights on English for the professionals' academic writing (research paper) purposes.

II. LITERATURE REVIEW

In retrospect of the previous research penetrating to linking adverbials, definition-similarities are notable in regard to their cohesion-building function. Halliday and Hasan term them as linguistic features that avail to the text unity, not only are they cohesion promoters but sequence markers and causality signifiers as well [15]. In concert with this, Biber *et al.* define LAs as critical cohesion-constructing linguistic devices [7]. On the one hand, LAs possess an effective function in elucidating the relation between two sentential units, at the same time, they reinforce the explicitness by manifesting the connection among text passages. It is also pointed out that, syntactically speaking, LAs are relatively trivial in the principal clause as they seldom make a difference in the overall clause structure, yet their main function is to make the logical link of clauses clear [1], such an objective and all-round interpretation makes it chosen as the theoretical definition of this study. But having said that, other scholars' definitions of LAs are equally significant. In Liu's point of view, LAs are powerful grammatical instruments for promoting cohesion in writing and oral register, which will benefit second language using, teaching, and learning [9]. Peacock affirms that LAs are not only a fit for uplifting context cohesion but are conducive to producing and highlighting claims [12]. In general, there might be diverse ways of comprehending LAs, but their syntactic function in facilitating discourse cohesion is commonly acknowledged.

Ways of taxonomizing LAs also vary among scholars. According to Biber and Gray [1], LAs can be classified into six groups from a semantic perspective, and they are respectively enumeration and addition, summation, apposition, result/inference, contrast/concession, and transition. However, in Halliday and Hasan's research [15], LAs are segmented into four classes: additive, adversative,

causal, and temporal [15]; which can be termed as the foundation of Larsen-Freeman and Celce-Murcia [16] classification: additive (emphatic, appositional, comparative), adversative (proper adversative, contrastive, correction, and dismissal), causal (general causal, causal conditional) and sequential [17]. This rationale has been employed by Liu in his cross-registered LAs research [9], which encompasses 110 LAs in sum and later proved efficient in many other studies [11, 14]. It is worthy of remark that this categorization is advantageous in being functionally systematic and specific. Interestingly, there is also a study that generates a taxonomized framework where LAs are mapped into 21 types (adversative, addition, listing, exemplification, condition, logical consequence, result, restatement, explanation, alternative, similarity, transition, comparison, opposite, summation, conclusion, signal of evidence, formulaic ending, initiating a lecture/program/conversation, marker of continuation, signal of returning to the main thread) after a semantic and pragmatic investigation [10], and such framework is essentially the integration of the seven classifications from four linguists [1, 17–19]. LAs research about English academic writing has been under the spotlight with diversified foci. Studying LAs from the perspective of interdisciplinary usage is a major heat. To cite Biber and Gray's academic register-specific research as an instance, profound scrutinization has been extended to a set of linguistic features' historical shifts [1]. Among these linguistic features, LAs are studied from both disciplinary and diachronic perspectives, being statistically uncovered that they are especially preferred by writers majoring in humanities and social science, and what is fascinating enough is that their usage in academic writing has gone through a salient decrease over the three centuries (1725–2005). Apart from that, there is a corpus-based study that compares the similarities and distinctions of LAs use in research papers of four disciplines (physics, computer science, linguistics, and management) by Chinese and native scholars [11], eliciting the findings that LAs usage is discipline-impacted, specifically, LAs occurs more frequently in soft non-science subjects than in hard-science subjects, which largely conforms to Biber and Gray's [1] assertion. Also, Chinese scholars are found to not only underuse additive and adversative LAs but show distinctive usage manners from native writers. However, only four LAs (*of course*, *yet*, *meanwhile*, and *therefore*) receive a qualitative interpretation, the way other LAs are improperly utilized in Chinese corpus is left veiled. Agreeing with Biber *et al.*'s [7] confirmations as to LAs' principal functions, Peacock obtains some new discourse outcomes brought by LAs in research papers: contrast or concession LAs are competent to achieve the claim-producing outcome, apposition LAs can achieve the claim-fortifying goal, and lengthy-clause-appearing LAs are able to make achieved the claim-approving purpose [12], which he calls a complement to Biber *et al.*'s [7] theories. Additionally, going beyond the previous finding that LAs emerge far less in scientific fields than in non-scientific ones, Peacock finds that to achieve the aforementioned goals, noticeable disparities lie in the LAs usage among heterogeneous disciplines of the same category (scientific or non-scientific).

LAs usage in a variety of registers is also a magnet to linguistic researchers. Biber and Conrad [20] conclude that LAs are “very common” in academic prose register, it is therefore justifiable to infer that LAs are frequently used in research article register since it is a sub-register of academic prose [1]. Holding the faith that painstaking register-based classification of LAs is of tremendous value in second language research and pedagogy, especially in academic English, a scholar produces a new taxonomy of LAs in view of semantics and pragmatics [10], confirming the preceding identification that LAs are crucial cohesion instruments in conversation and academic registers (academic written prose and academic lecture) and providing a novel modal for LAs analysis. Another sound corpus-based LAs analysis is also carried out on the basis of different registers [10], whose author inventively canvasses 110 LAs' usage patterns and times in five BNC registers, drawing the assertion that some LAs variants can be emphasized or neglected in specific English teaching registers according to their significantly higher or lower occurring frequency (sequential and summative LAs for journalism English education, for example). These constructive research makes tested that certain categories of LAs frequently occur in some general academic registers, while cases in some sub-registers like academic conference report, research paper, or dissertation have yet to be investigated, which is one of the motivations of the current study.

How LAs are used by learners in different academic stages has always been captivating, most of these studies are corpus-based and comparative in essence. Early in the last century, LAs were assured to be logical connectors that symbolize higher-level writing competence in Hong Kong English as a Second Language (ESL) freshmen's essays [8], students in which research are found to overuse adversative LAs *however* and *nevertheless* and misuse *on the contrary* when trying to signify contrast relationships, based on these finding, Crewe calls for more awareness-raising instructions of LAs usage for ESL writers. Similarly, targeting first-year undergraduates in Korea, Ha [21] conducts a comparison of the LAs usage frequency and patterns in their argumentative writing versus American students' writing, making the conclusion that Korean students are inclined to overuse all types of LAs, especially the types of sequential and additive, even though a similar LAs distributing pattern of two corpora is presented. However, there is a limitation that qualitative analysis is confined to a small part of LAs: *moreover*, *so* and *therefore*, which inspires the current study to conduct a well-rounded qualitative interpretation. Same in ways of selecting the participant groups (undergraduates) and the corpora register (argumentative essays), Appel and Szeib [13] are ingeniously oriented to participants' varied L1 backgrounds (Arabic, Chinese, and French), disclosing that L1 background does exert an influence upon LAs usage in English academic writing and that a certain number of grammatical issues could be ameliorated by Data-Driven-Learning (DDL hereafter), some pedagogical suggestion like integrating the corpus with LAs teaching are consequently come up with. Speaking of DDL in LAs-related study, it is necessary to mention another fine-designed research [15], which resembles Appel and Szeib's work in the part of comparing the LAs usage by

students from five distinctive L1 environments, looking into the effectiveness of corpus-mingled-DDL approach in the instruction of academic writing-based LAs usage [13]. In fact, back in a decade ago, there was analogous work that had to do with how concrete pedagogical approaches work on students' LAs use—It is unconcealed by Leedham and Cai [19] that Chinese students excessively use LAs such as *besides* or *on the other hand*, and that they have a preference for initial-positioning LAs, the authors hypothetically owe these linguistic issues to the Chinese teaching materials and sample answers that students were exposed to before studying in UK [18]. In terms of how LAs are used by English learners at an advanced academic stage, Chinese linguistics Ph.D. students' dissertations are compared with professionals' published research papers [14], from which Chinese doctoral students are observed to overuse and underuse dozens of LAs with hinging on some specific LAs in a fixed way. There is a surprising accord that these learners misuse *besides* and *actually* as the undergraduates do in Leedham and Cai's research [19], nonetheless they are in a relatively high academic phrase and specializing in linguistics. Well-devised and fine-analyzed though these two studies are, they were published around one decade ago, making it hard not to wonder if any historical changes in LAs usage during these ten years will generate different findings, which is also a part of the reason that the present study is conceived.

As is reviewed, past LAs studies are carried out from various perspectives consisting of but not confined to cross-discipline, cross-register, and pedagogy. Though most of these studies' targeted participants, scenarios, or findings are inextricably linked or consonant, research directly delving into LAs usage in Chinese linguistic-professional writers' published articles remains uncharted. Sophisticated and skillful as they are in exerting linguistic features, there must be more room to polish the use of LAs by correcting some improper use, thereby promoting the overall cohesion and quality of the paper. Inspired by the preceding research findings and gaps with realizing the significance of studying the LAs usage by Chinese linguists, this study endeavors to conduct a corpus-based study to find some overuse, underuse, or misuse of LAs based on a statistical comparison with English native linguists. To scratch beneath the surface of statistical tests for an in-depth linguistics study, a fine-grained qualitative analysis will be implemented. Analyzing in view of both taxonomized and specific LAs, this corpus-based comparative study is driven by addressing the following two questions:

Q1: Are there any differences between the LAs usage frequency of Chinese professionals' papers and that of native professionals' papers? If so, to what extent do such differences exist?

Q2: Based on the statistical result of Q1, what LAs are overused, underused, or misused by Chinese professional writers?

III. CORPUS AND METHODS

A. Corpus

The compiled corpus comprises two sub-corpora (CH-Corpus and EN-Corpus), which respectively include

published linguistics articles written by Chinese scholars and native scholars. The chosen linguistics articles are excerpted from 18 high-impacted and peer-reviewed international linguistic journals (Q1) according to the impact rank provided by JCR so that the quality and representativeness of the texts could be in non-doubt.

To ensure the excerpted texts match the research questions, Pan *et al.*'s [22] ground rules for text-selecting are referenced in this study. First of all, instead of being matched for the text number, the corpora will be matched for the token numbers. In addition, both the affiliation location and the full name are taken into account when selecting texts composed by Chinese or native writers. That is, only research papers with both Chinese full-name authors and affiliations of mainland China are accounted into CH-Corpus, and those of full names and affiliations in English-speaking regions, are elected to EN-corpus.

Concerning the text cleaning, some parts of the rigorous procedures innovated by Sun and Lan [23] are followed. In particular, research articles downloaded in PDF form experience a transition into plain text form via AntFileConverter [24], following which, trivial text sections containing meta information, abstract, reference, notes, acknowledgment, tables or figures, non-English characters, and supporting materials are manually excluded from both corpora. It is noteworthy that non-English characters emerging in the reserved parts are substituted by @, and that the extracted dialogues not written by authors are laboriously sorted out and discarded. In the interest of a well-organized text format, EmEditor [24] is a troubleshooter here for blank lines and redundant spaces. All these steps shape a well-constructed large corpus inclusive of two sub-corpora with a total token number of 1.03 million (513,001 tokens for the CH-Corpus, 508, 120 tokens for the EN-Corpus), which paves the way for the further comparison and interpretation.

B. Linguistic Framework

The linguistic framework in this study is a four-category LA list adapted from Liu's 110 LAs list: Absorbing the primary parts of several grammar textbooks, he develops an all-embracing LAs summary, where there are 110 LAs divided into four functional categories (Additive, Adversative, Causal/Resultative, Sequential). Plus, this list is practice and history-proof due to its broad adoption by academics [11, 13, 14, 26]. All these make it a perfection for the theoretical reference and basis of the current study. However, this study is more of a specific register (research article sub-register) study than a cross-register one, which means some adaptations need to be made through a pilot test so that this LAs framework could be research question-oriented and register-tailored.

C. Analysis Procedures

Based on the above-mentioned corpora and linguistics framework, four concrete procedures following a pilot test are designed for the analysis of two research questions.

Pilot Test: Concordance the 110 LAs in two sub-corpora using the N-Gram and KWIC function of Antconc (version 4.2.0) with a frequency threshold of 20 times, after which a 57-word LAs list is put in place.

Procedure 1: Prior to the calculation of raw frequency and

normalized frequencies (frequency per 10,000 words) of the adapted LAs list, for the good of analyzing accuracy, a qualitative analysis is implemented by cross-checking all the concordance lines to delete some cases that the concordanced chunks are not serving as LAs.

Procedure 2: From a macro-linguistic point of view, the normalized frequencies in view of the functional classes are presented, and a Pearson Chi-square test ($p < 0.05$) with the application of SPSS 27 is arranged for comparing the frequency-difference in LAs categories.

Procedure 3: From a micro-linguistic point of view, the normalized frequencies of every LAs are given. In like manner, a Pearson Chi-square test ($p < 0.05$) is done for a comparison of specific LAs' usage frequency.

Procedure 4: Based on the statistical results, by extracting and interpreting some typical instances, a follow-up qualitative analysis is provided to elaborate on the overused, underused, and misused LAs by Chinese professionals.

IV. RESULT AND DISCUSSION

A well-presented results section coupled with a convincing discussion will definitely prove the novelty and importance of your study. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

A. Frequency and Pearson Chi-Square Analysis of the LAs

Q1: Are there any differences between the LAs usage frequency of Chinese professionals' papers and that of native professionals' papers? If so, to what extent do such differences exist?

To answer the first question, the results of frequency and Pearson Chi-square analysis are yielded in tables. Macro-linguistically speaking, there is a significant difference with a 0.019 p -value in the sum amount of LAs between two corpora, which signifies that native writers attach greater importance to making use of LAs in the cohesion-building of research articles (Table 1). Among these four functional categories, adversative and causal/resultative LAs classes are prominent in their significant differences with p values of 0.000 and 0.041 (Table 2), indicating that Chinese linguistic professionals have yet to thoroughly achieve the goals of adversative or concessive, contrastive, correction, dismissal, general causal and conditional causal LAs as natives do [9].

This finding concurs with the discovery of Liu's and Biber *et al.*'s works to a large extent, where the former three categories of LAs are characterized with prominently high occurring frequency in academic prose and the case of sequential category is on the contrary [7, 9]. However, a contradiction is detected: the additive category of LAs is used most often in the general academic register in these two studies [7, 9], yet in the current study, it is the adversative category that is mostly employed by English natives. The reason might spring from the nuances of sub-registers' situational characteristics determined by different audiences and communicative purposes. That is to say, as one of the sub-registers of the academic register, the ultimate goal of the research article is either making new research contributions or making other experts convinced of

the research value, which thereby induces the preference for specific linguistic devices [20]. To be more concrete, far from imparting the already constructed knowledge through the discourse of introducing, depicting, and characterizing qualities like textbooks or teaching materials do, the principal communicative purpose of the research article is to convey critical and innovative thinking that requires idea collision, assertion push-over, and arguments or rebuttals. That explains why there are more adversative LAs with concessive, contrastive, and dismissal functions occurring in the research-article sub-register, and why the frequency of adversative LAs precedes that of additive LAs by English natives in this research.

The result is also in harmony with another LAs research in the part that the higher norm frequency of total LAs is discovered in English natives' corpus [11], but inconsistent with the part that all LAs categories in the latter study keep a notably higher occurring frequency in natives' corpus, to explain which, a detail of the corpora is tweezered out: the corpora in Gao's work are compiled with texts from 2000 to 2014, whereas the corpora in this study absorb the texts of the latest eight years [11]. Thus, it is entitled to conjecture that Chinese scholars have made substantive improvements in exerting some categories of LAs as the years go by. Similarly, such grammatical enhancement can also be acquired with the proceeding of learners' academic levels [1, 22], which is even more salient when referring to another LAs study [14], where not only is the frequency of total LAs number but also that of three LAs categories are much higher in Chinese doctoral students' corpus, with a considerably lower frequency of the adversative-type LAs. Such linguistic phenomenon is interpreted by the author as the coexistence of overuse and underuse by Chinese linguistics-major students, and it is plausible to speculate that some linguistics melioration of the LAs imperfectness has been received in the transition of student-to-researcher, thus the occurring frequencies of specific LAs types in CH-Corpus (additive, sequential) has been akin to those in EN-Corpus, though some significant differences are still in existence.

To discuss from the view of micro-linguistics, the results of LAs frequencies and the corresponding difference test are visualized in Table 3 and Table 4 (Complete version please see the appendix). In Table 4, *however* stands out due to its highest occurring frequency in both EN and CH corpus, respectively up to 696 and 573 and many times as the frequencies of *besides*, *second*, *again*, and *meanwhile*. Besides that, *for example* as well as *that is* are the second and third most common LAs in two corpora, with raw frequencies of 378 versus 324 and 315 versus 264. These data buttress Biber and Conrad's discovery that these LAs are commonly used in academic prose [20].

On top of that, huge disparities in usage times do exist among the calculated LAs, though some of whom share a certain resemblance in morphology or semantics: *in addition* and *in addition to*, *second* and *secondly*, *in/by comparison* and *in/by contrast*, for instance. These linguistic phenomena will be elucidated in the following qualitative analysis through delving into the concordance lines.

When it comes to the distinctions of usage frequency in Table 3, there are 32 out of 57 LAs bearing salient

differences, and 14 of them even show extraordinarily evident significances ($p = 0.000$), signifying that Chinese linguistics scholars have some defect in producing fully cohesive discourse in research articles with native-like application of LAs. As is pointed out by Biber and Gray [1], the primary discourse function of LAs is to clarify the logical relationships among clauses for the benefit of

context explicitness, which is also assured as a strong impetus of better cohesion in academic proeses [2]. In this case, the overuse, underuse, and misuse of these LAs by Chinese writers merit an in-depth qualitative analysis while having specific text instances included, and that will be carried out in Question 2.

Table 1. Frequency of LAs functional categories

Items	Additive	Adversative	Causal/Resultative	Sequential	In Total
Chi-square (a)	0.002	32.053	4.174	0.000	5.4914
In Total	1	1	1	1	14
Asymp. Sig.	0.963	0.000*	0.041*	0.995	0.19*
Effect Size (Cramer's V)	0.000	0.006	0.002	0.000	0.002

Table 2. Pearson Chi-square output of LAs functional categories

LAs Categories	EN-Copus Raw Frequency	Norm Frequency	CH-Corpus Raw Frequency	Norm Frequency
Additive	1748	34.4	1762	34.35
Adversative	1889	37.18	1573	30.66
Causal/Resultative	1006	19.8	1110	21.64
Sequential	805	15.84	813	15.85
In Total	5448	107.22	5258	102.49

Table 3. Frequency of all LAs

LAs	EN-Corpus Raw Frequency	Norm Frequency	CH-Corpus Raw Frequency	Norm Frequency
additionally	58	1.14	88	1.72
also (sentence initial)	36	0.71	25	0.49
also (in and also)	39	0.77	22	0.43
also (in but also)	90	1.77	122	2.37
again (sentence iinitial)	30	0.59	6	0.12
as well	46	0.91	26	0.51
besides	6	0.12	40	0.78
in addition	161	3.17	141	2.75
in addition to	74	1.46	79	1.54
furthermore	92	1.81	82	1.6
moreover	57	1.12	111	2.16
that is	315	6.2	264	5.15
in other words	46	0.91	88	1.72
for example	378	7.44	324	6.32
for instance	129	2.54	149	2.9
namely	50	0.99	105	2.05
likewise	25	0.49	23	0.45
similarly	116	2.28	67	1.31
at the same time	16	0.31	33	0.65
however	696	13.7	573	11.17
of course	33	0.65	19	0.37
nonetheless	25	0.49	19	0.37
nevertheless	52	1.02	56	1.09
though	233	4.59	112	2.18
yet (sentence initial)	16	0.31	31	0.64
yet (in other positions)	70	1.38	40	0.78
actually	36	0.71	48	0.94
in/by contrast	137	2.7	121	2.36
in fact	76	1.5	47	0.92
on the other hand	84	1.65	82	1.6
instead	96	1.9	43	0.84
rather	96	1.9	48	0.94
despite	76	1.5	109	2.12
still	129	2.45	162	3.16
accordingly	15	0.3	50	0.97
as aresult (with of)	19	0.3	21	0.41
as aresulty (without of)	20	0.39	48	0.94
because of	61	1.2	55	1.07
consequently	35	0.69	34	0.66
hence	43	0.85	118	2.3
so (sentence initial and after a comma)	92	1.82	128	2.5
Therefore	247	4.85	290	5.65
Thus	422	8.3	338	6.59
Otherwise	52	1.02	28	0.55
First	135	2.66	151	2.94
Second	88	1.73	99	1.93
secondly	7	0.14	21	0.41
third	22	0.43	35	0.68
finally (sentence initial)	149	2.94	55	1.07
lastly	22	0.43	21	0.41
next	104	2.05	37	0.72
then (sentence initial and after a comma)	151	2.97	135	2.63
then (and then)	86	1.69	156	3.04
at the same time	16	0.31	33	0.64
meanwhile	7	0.14	49	0.96
in summary	18	0.35	21	0.41

Table 4. Pearson Chi-square output of LAs with significant differences

LAs	Chi-square (a)	df	Asymp. Sig.	Effect Size (Cramer's V)
accordingly	18.515	1	0.000	0.004
additionally	5.882	1	0.015	0.002
again (sentence initial)	16.231	1	0.000	0.004
also (in and also)	4.902	1	0.027	0.002
as a result (without of)	11.264	1	0.001	0.003
as well	5.479	1	0.016	0.002
at the same time	5.737	1	0.017	0.002
besides	24.808	1	0.000	0.005
despite	5.576	1	0.018	0.002
finally (sentence initial)	44.227	1	0.000	0.007
for example	4.689	1	0.030	0.002
hence	34.231	1	0.000	0.006
however	13.144	1	0.000	0.004
in fact	7.118	1	0.008	0.003
in other words	12.768	1	0.000	0.004
instead	20.722	1	0.000	0.005
meanwhile	31.102	1	0.000	0.006
moreover	16.848	1	0.000	0.004
namely	18.997	1	0.000	0.004
next	32.486	1	0.000	0.006
of course	3.905	1	0.048	0.002
otherwise	7.432	1	0.006	0.003
rather	16.465	1	0.000	0.004
secondly	6.867	1	0.009	0.003
similarly	13.596	1	0.000	0.004
so (sentence initial and after a comma)	5.553	1	0.018	0.002
that is	4.996	1	0.025	0.002
then (and then)	19.589	1	0.000	0.004
though	43.618	1	0.000	0.007
thus	10.112	1	0.001	0.003
yet (sentence initial)	4.645	1	0.031	0.002
yet (in other positions)	8.472	1	0.004	0.003

B. Qualitative Analysis of the LAs

Q2: Based on the statistical results of Q1, what LAs are overused, underused, or misused by Chinese professional writers?

Before elaborating on this question, standards of overuse, underuse and misuse should be put on the table. For overuse and underuse, some comparative study is assigned with a 1.5× frequency of the interested corpora [21], while others define a 10-occurring time difference or a top 10 LAs-per category as a cut-off [11, 14]. In a word, these standards are arbitrary [27], though they are quantitative in essence, based on which, a 0.05 *p*-value of the difference test is determined as a threshold here. For underuse, its definition varies among works as well. Sometimes misuse is counted as both a reason and a result of overuse with salient mistakes [21], also, it is reckoned to be inaccurate-use in semantics [11]. In the present study, since the article writers are linguistics professionals with advanced language proficiency, misuse here is perceived as non-native use while acknowledging its coexistence with overuse, it is also believed that misuse can also co-exist with underuse, which will be exemplified by text samples.

For the purpose of meticulous interpretive analysis, only LAs with a particularly prominent *p*-value (*p* = 0.000) are discussed, which are *again*, *besides*, *moreover*, *in other words*, *namely*, *similarly* (additive), *however*, *though*, *instead*, *rather* (adversative), *accordingly*, *hence* (causal/resultative), *finally*, *next*, *then* (and then), *meanwhile* (sequential). The discussion is progressed in a clustered way on the basis of Liu's [9] four LA categories.

1) Additive category

Generally speaking, Chinese linguists are capable of

applying all the additive LAs in a native-like way. However, overuse in this LA category is obvious, which involves *besides*, *moreover*, *in other words*, and *namely*, plus, the underuse of *again* and *similarly* is also unignorable. These grammatical issues indicate that some LAs' communicative purposes of supplementing and paraphrasing messages are overestimated, while that of placing emphasis and listing similarities are underestimated.

For the LA *again* (0.59 versus 0.12) used in the sentence-initial, in both extracted sentences, it is used for stressing some point or fact. As for the reason for its low occurring frequency in CH-Corpus, it might be because that most Chinese writers offer it a sentence middle or end place as a multi-element modifier or circumstance adverbial [7], mostly applying it to modify the verb with conveying some repeated meaning.

Text instance 1

EN-Corpus: **Again**, this leaves unclear how much stock to put in MCDI data for predicting expected real-time word comprehension for a specific set of items.

CH-Corpus: **Again**, multiple regression analysis would allow us to alleviate estimation bias by simultaneously considering the effects of various factors.

For *similarly* (2.28 versus 1.31), according to the sample instance, its only location is at the beginning of the sentence as an additive server [12]. As an LA indicating that some new discourse is about to be added right after the former ones, it normally takes a certain degree of homogeneity with the previous information in semantics [7]. The far more lower usage frequency reflects Chinese scholars' underuse issue, this LA's effectiveness in stringing the alike information in a cohesive way is not fully realized.

Text instance 2

EN-Corpus: **Similarly**, we find evidence against the hypothesis that learners with lower aptitude at pre-test improve more in LV.

CH-Corpus: **Similarly**, in the second conversation, David answered Shawn's question with "no issue", when he meant "no problem".

For *besides* (0.12 versus 0.78), a common additive-belonging LA for message adding and juxtaposition or progression specifying, is of weirdly scarce application in native research paper context. Both native and Chinese scholars are inclined to put it at the initial of sentences for such purposes nonetheless, the great disparity of frequency indicates that Chinese scholars overtly use this LA like many other Asians do, no matter their linguistic and academic level [14, 18, 20].

Text instance 3

EN-Corpus: **Besides**, sometimes children can have a bad peer group which will effect on the child.

CH-Corpus: **Besides**, spoken registers have higher mean dimension scores than written registers or meta discourse occur more frequently in in speech than in writing.

For *moreover* (1.12 versus 2.16), according to the text sample, both groups apply this LA to mark extra or progressive information at the sentence-initial, which also indicates a stronger tone than *besides*. However, the nearly twice difference in frequency also suggests Chinese linguistic scholars' overuse of this LA, which is also an issue among linguistic-specializing doctoral students who own a Turkish L1 background [2].

Text instance 4

EN-Corpus: **Moreover**, the fact that advanced L2 learners produce fluent and sophisticated runs is no guarantee that these runs are PUs.

CH-Corpus: **Moreover**, the completion of challenging tasks would enhance student self-efficacy.

Syntactically and semantically speaking, *in other words* (0.91 versus 1.72) and *namely* (0.99 versus 2.05) do have qualities in common, both of whom belong to the additive taxonomy for a further explanation function and refer to conveying the same thing in the other way, that is why they are put together for discussion here. Clearly, the former is situated at the sentence-initial, while the latter is located at the middle place for reflecting the above-mentioned discourse purposes in both corpora, signifying that Chinese professions are capable of using these two LA in a native manner. Yet, the nearly doubled usage frequency of CH-corpus brings the problem to the surface that Chinese professionals rely too much upon these two LAs. Going back to the frequency list makes it obvious that the LA *that is* showing up 315 times in EN-Corpus is much more commonly used in native research article context for the same communicative functions, which is an instance that deserves learning in research article writing.

Text instance 5

EN-Corpus: **In other words**, some learners seem better able than others to make the most of what skill and knowledge they have.

CH-Corpus: **In other words**, the multilingual self system develops from the self-organizing dynamic interaction of international components and external forces in a spatiotemporal dimension.

EN-Corpus: we have not found any ergative language that exemplifies our key prediction mentioned above, **namely**, one in which switch reference tracks (absolutive) intransitive subjects and objects, to the exclusion of (ergative) transitive subjects.

CH-Corpus: The model was first applied to the data in a pilot study that was guided by the objective of the analysis, **namely**, to describe and, subsequently, to compare the discourse features of HRT and LRT during classroom interaction.

With these real texts and interpretation, it can be inferred that hinging on additive LAs excessively is possible to weaken the conciseness of the research paper context, for the reason that the more the LAs are, the more sentences there would be. Additionally, the ignorance of *again* and *similarly* might decrease the clarity when presenting some emphasis and likeness. More fascinatingly, most of the analyzed additive LAs are arranged at sentence-initial, and such a sentential position makes the statement discourse much more organized and systematized.

2) *Adversative category*

Unlike the usage situation of additive LAs, Chinese professionals' primary issues of adversative LAs are underuse and misuse, which is exemplified by data and text instances. Misuse is of particular prominence in this category, which happens to almost all the discussed LAs other than *rather*. Concretely, ignorance of the medial sentence position of *however* and *instead* is reflected in CH-Corpus, and the initial sentence position of *though* has not received enough attention either. Further, Chinese authors have a collocation issue in *though* and *instead* by employing these LAs with other adverbs or prepositions, which is not the way the English natives do. As for underuse, an around two-time difference in usage frequency exists in *though*, *instead*, and *however*, unmasking Chinese authors' weakness in using adversative LAs for the discourse purpose of carrying out the rebuttal and argument in a cohesive way.

For *however* (13.7 versus 11.17), serving as a cohesion contributor by highlighting the adversative relation, its significance in the academic writing register has been supported in a variety of research [7, 11, 14, 28]. Sufficiently interesting, *however* is of the highest frequency in both corpora of this study, yet there is still a significant difference of its frequency between two corpora. In reviewing other studies, the underuse of *however* is found to be a shared problem among L2 English learners with distinctive L1 backgrounds and linguistic competence [11, 13, 14, 25, 28]. In this study, a contributing factor is surmised through digging into the concordance lines, that is, apart from using *however* as a sentence-initial adverbial, English native writers also provide it with a sentence-medial position, which is still awaiting the recognizance of Chinese writers who prefer to put this LA at sentence beginning, this linguistic phenomenon aligns with Biber *et al.*'s finding that it is ubiquitous of a sentence middle-positioned LA like

however emerges in the academic register [7]. The preference for sentence-initial *however* also lies among Chinese undergraduates based in UK universities, nevertheless these chosen participants are of a proficient linguistic level [19].

Text instance 6

EN-Corpus: Research on the typological distribution of person systems, **however**, has found evidence that only four of these are grammaticalized as person categories.

However, L1 writers seem to be shifting their writing from more explicit expressions of stance (I think) to more implicit expressions of stance (the fact that...; the problem is that...).

CH-Corpus: **However**, two infinites embedded in a conditional cannot co-refer, as shown in (i).

For the LA though (4.59 versus 2.18), it is an inter-clauses contrast marker, which thus belongs to the adversative/concessive category. Even though it is proven to be used more often in the speaking register instead of the academic register [7, 9], odd enough, not only is its usage frequency in EN-Corpus above 200 times, but its normalized frequency is almost twice as that in CN-Corpus, which requires more in-depth investigation in future study. Regarding the manner of use, from the text sample it can be found that English natives are accustomed to setting it at the sentence initial to either elicit a clause or modify other adverbs or adjectives, Chinese writers, however, tend to give it a sentence medial position and make it co-occur with even.

Text instance 7

EN-Corpus: **Though** preliminary, the results of this section corroborate what is found with the by-adjuncts, in ways that could be extended in several directions.

Though the subject of the matrix clause is some female individual and that of the embedded clause is the speaker, the different subject suffix does not appear.

CH-Corpus: You cannot be aggressive or emotive, **even though** you really like or hate someone's work.

What is said above applies to the teaching of English rhetorical figures, **though** there have been studies of a concern with teaching or learning English rhetorical or figurative language.

For *instead* (1.90 versus 0.84), its raw frequency in two corpora (96 versus 43) is consistent with the statistical results in Liu [9] and Biber *et al.*'s research [7], this LA occurs less than 100 times in the academic register. Characterized with different communicative purposes by scholars for doing correcting or contrasting [7, 17], it also showcases divergent use manners in two corpora (See the text instance). Except for the medial-sentence place, *instead* takes place in front of a clause for a contrastive outcome in EN-Corpus, by contrast, it mostly happens at sentence-initial or end in Chinese linguists' excerpts to correct the preceding discourse. According to the frequency difference, it can be inferred that *however* is underused by Chinese professionals, which is a support of Granger and Tyson's report about its underuse by non-native English users [28].

Text instance 8

EN-Corpus: The increased time to produce the verb across conditions may **instead** indicate that participants are focusing on correctly repeating the preamble before they plan their verb completion as opposed to planning the full response prior to onset.

We argue **instead** that embedded clauses in this context are better understood as verbal modifiers.

CH-Corpus: As implied in Beavers, Levin, and Tham 2010, path information in motion event constructions is a descriptive and semantic label, **instead of** a syntactic one.

Instead, the teacher's implemented or even further developed the relevant language policy.

Rather (1.90 versus 0.94) is an LA subordinated to the adversative group with a correcting function as well, whose raw frequency in EN-Corpus is amazingly the same as *instead*. Here are some speculations for its low usage frequency: for one thing, the discourse function of correcting is less pervasive in the academic register, especially in the research-article register, for the other, seldom is it solely utilized but associated with other words. As the extracted sentences show, both English and Chinese writers normally collocate it with *than* at a medial-sentence position, which creates a more concise and cohesive context outcome.

Text instance 9

EN-Corpus: Automatic Versus Controlled Language Processing Frequency effects are more often evidenced in automatic online language processing **rather than** in more considered production tasks where there is more time for conscious creation and editing.

CH-Corpus: Compared to their previous EAP courses, they found that they were expected to explain their points in their written response to each question **rather than** using a variety of writing techniques (e.g., definition, comparison, contrast).

To conclude, the misuse and underuse of adversative LAs are highly possible to decrease the accuracy and criticalness of the research article, sometimes affecting the context cohesion and consistency, which is worth in-depth observation and emendation with the help of corresponding academic prose corpora.

3) Causal/Resultative category

In this LAs category, common imperfect usage in CH-Corpus includes overuse, misuse, and their co-existence. The following analyzed LAs (accordingly, hence) are used with an almost trinal frequency in comparison to the EN-Corpus. Besides that, Chinese professionals prefer a sentence-initial place, probably because they intend to remind the audience of the casual links between clauses. However, the manner native professionals achieve such communicative purpose is different: they put them at the sentence middle or end, thereby bringing more compact information and more explicit causality.

For *accordingly* (0.30 versus 0.97), it appears merely 15 times in EN-Corpus, which matches with the statistical results in Liu's and Peacock's studies [9, 12] where this LA is also of low usage frequency in academic prose, signifying that it is truly overused by Chinese professionals, which is also a problem among Turkish PhD students who

specialize in applied linguistics [26]. In regard to the usage style, native professionals arrange it more at the sentence end or middle than at the sentence-initial, rendering the logical relationship clear and the context information condensed.

Text instance 10

EN-Corpus: As this study and others have shown, students who have some experience with related genres may encounter fewer issues when attempting to apply or transfer that knowledge **accordingly**.

CH-Corpus: **Accordingly**, many language teaching researchers began to explore how to use the cooperative learning method to improve the result of second language learning.

For *Hence* (0.85 versus 2.30), as a word commonly serving to express general causality in the academic register [9], it makes discernible the results or consequences of the previous discourse [7]. From the occurring frequency almost three times higher, Chinese professionals are deduced to overuse this LA. From the text examples, Chinese professionals are noticed to usually utilize this LA independently with pitting it at the sentence initial. On the other hand, however, English natives tend to locate it at the sentence medial or use it with *and*, making the context more information-abundant by prolonging the sentence length, which is an effective cohesion-facilitating manner deserves learning from [14, 25].

Text instance 11

EN-Corpus: For example, at one end of the spectrum was the thriver group with very high self-efficacy and strategic self-regulation, low levels of anxiety, and **hence** high buoyancy levels; at the other the dependent group with the very lowest levels of self-efficacy and strategic self-regulation but high anxiety and low buoyancy.

CH-Corpus: **Hence**, the context in Kashack *et al.*'s study has little impact on the processing of the primes, and the cumulative effects they observed might largely rely on explicit memory.

As above discussed, the improper use of causal/resultative LAs seems to be widely confronted. Appropriate sentential positions and association with other conjunctions or adverbials make it possible to embrace more information, while keeping the context consistent at the same time, which is more than significant in academic writing.

4) *Sequential category*

In this category, the improperness of sequential LAs usage by Chinese professionals is threefold, say, underuse, overuse, and misuse. Similar to the above-mentioned case, some LAs are associated with other conjunctions at times, which in practice is not the prevalent way of enhancing cohesion by English native professionals. What is more, in view of the sentential position, some LAs usage ways by natives, which are employed for the sake of explicitness and cohesion, have not gained the full recognizance of Chinese professionals. Weakness in utilizing sequential LAs is likely to refrain the sequence and order from being definite and explicit, sometimes ambiguity and redundance-inviting.

Finally (sentence initial; 2.94 versus 1.07), a LA with functions of enumerate and list information, is confirmed to be frequently applied in academic proeses [9]. Through

observing the text sentences, it is safe to say that a sentence-initial position does make the numbering and sequencing discourse more distinct and that Chinese linguistic writers are able to apply this LA as English native writers do. However, the underuse issue is also salient according to the compared frequency. Attention-catchingly, such problem is also reported in Lei's linguistic Ph.D. students-targeted work, which, to some extent hints that the underuse of *finally* is independent of writers' disciplines and linguistic proficiency [14].

Text instance 12

EN-Corpus: **Finally**, we may consider the importance of playfulness (that is, a playful stance) in L2 learning.

CH-Corpus: **Finally**, based on our research findings, we summarize and highlight a number of potential implications for teaching and pedagogy.

The LAs *next* (2.05 versus 0.72) and *then* (and then; 1.69 versus 3.04) are semantically and functionally in common, which explains why they are clustered to be analyzed here. A part of Liu's finding regarding these two LAs is quote-worthy here [9], that is, nonetheless both of them bear the function of producing or numerating the message in sequence, and their used frequencies in academic register differ 18 times (7.26 versus 130.27). Back to the concordance lines, divergence in the using manner of *next* is exposed--it is respectively put at the sentence end and sentence beginning by natives and Chinese. As the sentence shows, a sentence-final position of *next* weakens the existence of LA, instead, it leads the audience to throw more focus on the conveyed information rather than the syntactic modifiers. As for *and then*, in this sample sentence it is applied to depict the experiment procedure in both corpora and its excessively high frequency in CH-Corpus possibly result from Chinese writers' overemphasis on stating the specific steps of the methodology, which might make the context less fluent and consistent.

Text instance 13

EN-Corpus: Corpus downsampling Each of the UKWAC, Wikipedia, and Reddit corpora were prohibitively large for analyses that we describe **next**.

CH-Corpus: **Next**, we asked two multi-lingual students (who did not participate in our study) to comment on the statements, and then we clarified possible confusion and eliminated redundancy.

EN-Corpus: In this section, we examine a particular by-adjunct modifier that produces the relevant contrasts, **and then** adduce additional diagnostics that achieve a similar effect.

CH-Corpus: (If I had a lot of money now) appeared on the slide, **and then** participants read it after the instructor.

Meanwhile (0.14 versus 0.96) is an argument-appealing LA specifying simultaneously happening incidents or newly inserted information [7, 9], the problematic use of which is prevailing among L2 writers. Not only are Chinese writers in linguistic major confronted with its overuse [14], but Chinese writers in other different disciplines and academic levels are perplexed by this problem [11, 19]. It is even detected by Gao [11] that Chinese research article composer misuse *meanwhile* instead of *furthermore*, and that they misuse it for a complete reference of things taking place at

the same time while native writers view it as more of an information-complement function LA. The latter linguistic phenomenon is also reflected in this study, moreover, Chinese linguistic writers are used to locate *meanwhile* at the middle of the sentence plus a co-occurrence with *and*, thus making the context less contact but unnecessarily lengthy.

EN-Corpus: **Meanwhile**, attempts to elicit metaphor-related skills/competences have been mixed. Instrumentation has varied in reliability, been limited in scope, and used arguably flawed reliability coefficients (McNeish 2018).

CH-Corpus: They were familiar with the CSE and contextualized instructional features, and **meanwhile**, they worked together with the students to mediate the national language standards in language instructional settings.

In summary, the complete comprehension of some sequential LAs' semantic meanings and syntactic roles is still a challenge to not only Chinese linguistics professionals but many other English L2 learners. To be exact, these discussed LAs are confused with time adverbials (*meanwhile*), or only supposed to be used as adverbs (*next*), sometimes their situational functions are overstressed (*and then*) or ignored (*finally, next*), all these give rise to the manifold imperfectness of the sequential LAs usage by Chinese professional writers.

Based on these tests and analysis, it can be concluded that Chinese linguistics professionals demonstrate imperfect use in all taxonomies of LAs, especially additive and adversative taxonomies. To be specific, apart from the misuse dotted in four LAs categories, Chinese professionals overuse most of the analyzed adversative LAs as well as all analyzed causal/resultative LAs, and underuse all the analyzed adversative LAs. These findings tally with Gao's LAs [11] research where Chinese scholars fail to use the same types of LAs as natives in multi-disciplinary research articles, yet there is a discordance with one of Lei's findings [14], which is, Chinese linguistics PhD students present overuse in a couple of adversative LAs that are underused by Chinese professionals. Concerning the comparison versus linguistic Ph.D. students with a Turkish L1 background, it is a coincidence that both Chinese and Turkish show the most imperfection in LAs of additive and adversative categories.

As mentioned before, additive and adversative LAs are the most frequently employed LAs types in academic context, their situational functions in stating the truth and clarifying the arguments are affirmative, the inappropriate usage of whom will possibly lead to some negative discourse outcomes. First and foremost, too much presence of LAs may result in more clausal lengthiness and less explicitness. Also, insufficient use of LAs refrains the discourse relationships from being more manifest and distinct. Last but not least, misusing LAs leads to misleading context meanings and less linguistic sophistication. All these play a counterproductive part in promoting the fluency, coherence, and accuracy of research article texts. As for the contributing factors of these issues, Ha's and Lei's viewpoint stands out first [14, 19], which is, it is the unabundant register, genre, and style awareness that results

in these LAs usage flaws, thereby decreasing the text coherence and the overall writing quality. Additionally, it is the insufficient reading of the native-made research articles that constrains Chinese linguistic professionals from familiarizing themselves with the flexible and authentic usage manner of these LAs. Mention-worthy still, some usage problems presumably stem from the non-English L1 background of writers, which is possibly irrelevant to their academic phase, specializing discipline and L2 proficiency.

There is also a necessity to realize that crucial as the LAs are in erecting cohesion, they are neither equal to text cohesion in a complete sense nor totally representing the text coherence and the overall writing quality [8]. The aim of this study is not to make it convinced that LAs are superior to any other grammatical features nor to advocate a preciosity on LAs usage in the academic writing register. Rather, it attempts to probe into the usage differences among writers with disparate L1 backgrounds and how these differences, to some degree, affect Chinese professional writers' text cohesion. In this circumstance, the results of this study ought to be viewed in a dialectical way—commitment and attention should be paid to perfect the LAs usage in English research article writing, yet other contributing factors of the text cohesion, coherence, and even the overall quality, noun phrases, and complement clauses, for instance, should not be ignored.

V. CONCLUSION

This corpus-based comparative study investigates the differences in LAs usage between the linguistics research papers written by Chinese and English native professionals. It is ascertained that in light of both usage frequency and manner, remarkable distinctions do exist in their research-article texts, an interpretive analysis also puts 16 problematically used LAs under the spotlight, observing that Chinese linguistic primarily overuse additive and causal/resultative LAs, underuse adversative LAs with an all-over-spread misuse. This study coincides with the work of Granger and Tyson [28] that inappropriate LAs use by L2 writers will, to a certain extent, impede the cohesion of their academic discourse. Also, this work is consistent with previous findings that Chinese academic writers show a tendency to excessively use additive LAs and underuse adversative LAs, though they keep dedicating to linguistics with a comparatively advanced comprehension of grammatical features [14]. The present study also is supportive of some textbook-works regarding the commonly occurring LAs in academic register and its sub-registers [7, 20]. Some new findings are unexpectedly obtained: misuse can be co-happening with overuse and underuse in every LAs taxonomy. Also, some LAs usage issues are commonly faced with L2 academic English writers with diverse biographic backgrounds.

Hopefully, the current study will generate some pedagogical implications on corpus-assisted English academic writing for professionals. First, as Lei emphasizes [14], register awareness-raising instruction should be carried out without further ado, with the enriched recognition of the grammatical variants' occurring position as well as frequency in specific registers and sub-registers, Chinese linguistics professionals will have more chances to

go beyond the level of error-free grammar usage to that of the native-like and accurate usage. Second, as opposed to a mechanical list of LAs, it is the corpus-based data driven learning embracing well-edited computational corpora that can be brought into full play in improving L2 writers' linguistic devices usage. A native and natural reading exposure can be offered with the help of authentic materials extracted from international linguistics journals [13, 28]. Third, register-specific collocation and position of LAs should be provided with due attention, which takes a critical role in boosting the cohesion and general quality of Chinese linguists' research article texts, where there has already been an advanced linguistic proficiency. Most importantly, grammatical analysis should be set as a top priority in the training and practice of English research paper writing, more specifically speaking, the integrated understanding of LAs' main classes, semantic categories, and syntactic functions is a must.

Several limitations of this study need to be acknowledged. Firstly, there is a certain probability the small size of the corpora leaves some LAs under-representative, namely, some LAs' low occurring frequency in this study may be produced by chance. What is also cast into doubt is that the small effect size of the probability value is exactly a consequence of the limited-sized corpora, further studies with larger corpora size are therefore advocated to increase the stability of effect size. Secondly, in consideration of Granger and Tyson 's assertion that LAs cannot completely determine the integrated cohesion [28], a follow-up correlation test to study the extent to which the use of LAs would affect the cohesion and writing quality of the research paper is a lack in this study, which awaits to be fixed later. Thirdly, distribution and dispersion of the grammatical features are also indispensable parameters in linguistic analysis [29], due to the limit of manual labor, however, they are not included in the methodology of this study, which deserves an implementation in the research ahead. Lastly, since the current corpora are compiled by research articles crossing a 10-year time span, chances are that some historical alterations in LAs usage overrepresent or misrepresent some grammatical issues. In this circumstance, the effect of the diachronic innovations of LAs use merits detailed scrutinization in proceeding research. In prospect of the future work concerning English research-article writing and EAP grammatical analysis, these limited sides are much expected to be amended and exploited. Providing alternative perspectives in both research and pedagogy, these limitations are not all duds. Rather, they make more forward-looking the research related to English academic writing (research article) for the professionals.

CONFLICT OF INTEREST

The author declares no conflict of interest.

REFERENCES

- [1] D. Biber and B. Gray, *Grammatical Complexity in Academic English*, Cambridge Core: Cambridge University Press, 2016.
- [2] G. Gilquin, S. Granger, and M. Paquot, "Learner corpora: The missing link in EAP pedagogy," *Journal of English for Academic Purposes*, no. 4, pp. 319–335, 2007.
- [3] G. Lan, Q. Zhang, K. Lucas, Y. Sun, and J. Gao, "A corpus-based investigation of noun phrase complexity in L1 and L2 English writings of a first-year composition course," *Journal of English for Specific Purposes*, vol. 67, pp. 4–17, 2022.
- [4] G. Lan and Y. Sun, "A corpus-based investigation of noun phrase complexity in the L2 writings of a first-year composition course," *Journal of English for Academic Purposes*, vol. 38, pp. 14–24, 2019.
- [5] X. Lu, "Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment," *Language Testing*, vol. 34, no. 4, pp. 493–511, 2017.
- [6] A. Zareva, " 'And so that was it': Linking adverbials in student academic presentations," *Relc. Journal*, vol. 42, no. 1, pp. 5–15, 2011.
- [7] D. Biber, S. Conrad, G. N. Leech, *Longman Student Grammar of Spoken and Written*, Pearson Education ESL, 2002.
- [8] P. L. Carrell. (1982). Cohesion is not coherence. [Online]. Available: <https://www.jstor.org/stable/3586466/>
- [9] D. Liu, "Linking adverbials: An across-register corpus study and its implications," *International Journal of Corpus Linguistics*, vol. 134, no. 4, pp. 491–518, 2008.
- [10] Z. Yin, "Register-specific meaning categorization of linking adverbials in English," *Journal of English for Academic Purposes*, vol. 22, pp. 1–8, 2016.
- [11] X. Gao, "A cross-disciplinary corpus-based study on English and Chinese native speakers' use of linking adverbials in academic writing," *Journal of English for Academic Purposes*, vol. 24, pp. 14–28, 2016.
- [12] M. Peacock, "Linking adverbials in research articles across eight disciplines," *Iberica*, vol. 20, no. 20, pp. 9–34, 2010.
- [13] R. Appel and A. Szeib, "Linking adverbials in L2 English academic writing: L1-related differences," *System*, no. 78, pp. 115–129, 2018.
- [14] L. Lei, "Linking adverbials in academic writing on applied linguistics by Chinese doctoral students," *Journal of English for Academic Purposes*, vol. 11, no. 3, pp. 267–275, 2012.
- [15] M. A. K. Halliday and R. Hasan, *Cohesion in English*, London: Longman, 1976.
- [16] D. Larsen-Freeman and M. Celce-Murcia, *The Grammar Book*, 3rd ed. America: Heinle ELT, 2016.
- [17] M. L. Walker, "Can data driven learning address L2 writers' habitual errors with English linking adverbials?" *System*, vol. 69, pp. 26–37, 2017.
- [18] R. Huddleston and G. K. Pullum, *The Cambridge Grammar of the English Language*, Britain: Cambridge University Press, 2002
- [19] M. Leedham and G. Cai, "Besides on the other hand: Using a corpus approach to explore the influence of teaching materials on Chinese students' use of linking adverbials," *Journal of Second Language Writing*, vol. 22, no. 4, pp. 374–389, 2013.
- [20] D. Biber and S. Conrad, *Register, Genre, and Style*, Cambridge Core: Cambridge University Press, 2009.
- [21] M. J. Ha, "Linking adverbials in first-year Korean university EFL learners' writing: A corpus-informed analysis," *Computer Assisted Language Learning*, no. 6, pp. 1090–1101, 2016.
- [22] F. Pan, R. Reppen, and D. Biber, "Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals," *Journal of English for Academic Purposes*, no. 21, pp. 60–71, 2016.
- [23] Y. Sun and G. Lan, "A bibliometric analysis on L2 writing in the first 20 years of the 21st century: Research impacts and research trends," *Journal of Second Language Writing*, no. 59, 2023.
- [24] L. Anthony. (2022). Antfileconverter homepage. Laurence Anthony's AntFileConverter. [Online]. Available: <https://www.laurenceanthony.net/software/antfileconverter/>
- [25] EmEditor Text Editor. (2023). [Online]. Available: <https://www.emeditor.com/text-editor-features/history/emeditor-free/>
- [26] H. Günes, "A corpus-based study of linking adverbials through contrastive analysis of L1/L2 PhD dissertations," *World Council for Curriculum and Instruction*, no. 2, pp. 21–38, 2017.
- [27] C. Chen, "The use of conjunctive adverbials in the academic papers of advanced Taiwanese EFL learner," *International Journal of Corpus Linguistics*, vol. 11, no. 1, pp. 113–130, 2006.
- [28] S. Granger and S. Tyson, "Connector usage in the English essay writing of native and non-native EFL speakers of English," *World Englishes*, no. 1, pp. 17–27, 1996.
- [29] V. Brezina, *A Practical Guide*, Britain: Cambridge University Press, 2018.

Copyright © 2024 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).