A Corpus-Based Study on the Semantic Use of Reporting Verbs in English Majors' Undergraduate Thesis Writing

Yueyue Huang

School of Foreign Languages, Guangzhou Xinhua University, Guangdong, China;
Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China

Abstract—This study aims to add further pedagogical knowledge on students' academic writing by investigating the semantic patterns of reporting verbs (RVs) in L1 Chinese undergraduate English majors' theses in a southern Chinese university based on the semantic categories by Hunston et al. (1996) and Charles (2006a). A comparative analysis was conducted across L2 and L1 students' academic writing in the discipline of applied linguistics. The study yielded two major findings: 1) there was a significantly insufficient employment of RVs in general, particularly among three categories (Argue, Show, Find) by L2 students, who also presented a strong reliance on argumentation by intuition; 2) L2 students illustrated a restricted vocabulary repertoire of colloquial RVs and their usage of RVs was misrepresented in context, diverging from the intended rhetorical functions. These findings indicate that evidence-based argumentative writing practice and targeted lexical and rhetorical instructions on vocabulary knowledge require further promotion in L2 English learners' academic writing training.

Index Terms—reporting verbs, semantic category, L2 undergraduate thesis, academic writing

I. INTRODUCTION

Undergraduate thesis, or Bachelor's dissertation, is a major academic writing task usually assigned during the final year of a Bachelor's degree program to examine students' critical thinking, rhetorical awareness as well as their disciplinary knowledge. Chinese students majoring in English are normally required to complete theses in their second language (i.e. English), following academic writing conventions, yet obstacles might occur when students are expected to shift from a general practice of English as a foreign language (EFL) to English for academic purposes (EAP).

To compose evidence-based argumentation, writers should exploit external sources to persuade target audience effectively, which is a complex, yet particularly crucial skill as it "situates the writer within a disciplinary framework" (Liardét & Black, 2019, p.37) and establishes his or her voice as a credible or "informed insider" (Hyland & Jiang, 2019, p.263). Successful academic arguers shall make their claims against a backdrop of existing perspectives, creating "a balance between introducing their own perspective, acknowledging the existence of other perspectives, and estimating what their audience's assumed perspective will be" (Miller et al., 2014, p.108).

Reporting verbs (RVs) are a key linguistic device that academic writers use to effectively synthesize and incorporate sources into their argumentation. It is a critical aspect of broader citation practices that student writers are to learn for evidence-based argumentation (Kwon et al., 2018). Inappropriate reporting practice marks the writer inexperienced and might result in spurious or tenuous argument pieces. Therefore, investigation on the employment of RVs in students' thesis writing can help teaching practitioners as well as students *per se* to better understand and respond to the challenges experienced by undergraduate thesis writers.

Relevant investigations have been pursued into the inclusion of RVs in academic writing mostly in terms of semantic categorization or evaluation (Kwon et al., 2018; Uba, 2019), syntactic patterns (Jarkovská & Kučírková, 2021; Shaw, 1992), types and functions by different populations (Liardét & Black, 2019; Marti et al., 2019; Thompson & Ye, 1991; Yeganeh & Boghayeri, 2015) as well as disciplinary differences (Hyland & Jiang, 2019; Jarkovská & Kučírková, 2021; Uba, 2019). All have provided valuable insights into the understanding of citation practice in scholarly writing. Among them, one central theme was the semantic and functional use of RVs, based on which taxonomies of RV have been proposed.

A. Semantic Categorization of RVs

One early attempt to categorize RVs is Thompson and Ye (1991), where RVs were investigated across disciplines from the introduction sections and two distinct categories were proposed: denotation (textual, research and mental) and evaluation (writer's stance, writer's interpretation, as well as author's stance). The three-fold evaluative potential of RVs was further pinpointed by Hyland (1999), in which the term "discourse" was rephrased by "textual" and "cognition" by "mental" RVs, and evaluation RVs were restructured as "factive (writer acceptance)", "non-factive (author's stance)" and "counter-factive (writer disagreement)". Nonetheless, the overlapping nature of Thomason and

Ye (1991)'s systems rendered detailed examination difficult to proceed, and the generalization from the mere introduction section seemed to be simplified. Additionally, as Hyland (1999)'s taxonomy emphasized the rhetoric function of RVs, such categorization might be more congenial for semantic evaluation.

More focused attempt on elucidation of semantic patterns of RVs was made by Hunston et al. (1996), which extended their exploration into a COBUILD project and identified four semantic categories of RVs: Argue, Find, Show, and Think. Charles (2006a) explained this categorization as follows:

1. ARGUE verbs (*argue, suggest, propose*, etc.) are concerned with writing and other forms of communication. e.g. *Hofstede argues that Value is the general tendency of people to like one rather than another ....*

2. THINK verbs (*think, assume, feel,* etc.) are related to thinking, including having a belief; knowing or understanding; hoping or fearing. e.g. *...the scholar assumes that children's literature is an important part of the literary Polysystem...*

3. SHOW verbs (*show, demonstrate, reveal,* etc.) indicate a fact or situation. e.g. *The findings above demonstrate that Explicitation do abound through the text.*

4. FIND verbs (find, observe, discover, etc.) are concerned with coming to know or think something. e.g. *From the above narrative, we can find that accuracy is very important...*

This categorization has been commonly adopted to examine the pattens of RVs used by different populations, mostly by L1 English speakers. Charles (2006b) found that RVs from the Argue category were more frequently used by native English MA thesis writers than the Find and Show verbs across contrasting disciplines of politics (social science) and material science (natural science). Friginal (2013) compared the L1 upper-level college writers and the professional writers in forestry and found that students used more Show and Think verbs and fewer Argue verbs than professionals did. Marti et al. (2019) set out to explore differences by levels of expertise (expert/novice) and nativeness (L1 English/Turkish) in applied linguistics, and reported that both native and non-native expert writers show little variation in their reporting practices yet remarkable variation ("discursively hybrid") was found between non-native novice writers and the other groups (p.98). These studies pointed out several key variables in RV usage patterns, including disciplinary diversity, levels of expertise as well as native/non-native disagreement.

However, there are relatively fewer studies on the RV patterns by EFL learners, especially amongst L2 English undergraduate learners. One investigation by Kwon et al. (2018)'s did point out that L2 undergraduate writers across majors in a first-year writing program had a general preference for Argue verbs among the four categories in order by Think, Show, and Find. But it is unclear whether such pattern will replicate itself in different disciplines. Other studies, though not focused on the patterns of RVs, have reported a more restricted set of verbs (Hinkel, 2003; Liardét & Black, 2019; Ramoroka, 2014) and preferred employment of conversational verbs (Granger & Paquot, 2009) in the reporting practice of English L2 writers compared with L1 professional writers, though learners deployed RVs with similar variability as their L1 classmates (Liardét & Black, 2019). Nonetheless, to provide more pedagogical suggestions, it is still in need of a further inquiry into the patterns of L2 undergraduates' citation practice.

B. Research Questions

The present paper thus aims to contribute to the pedagogical knowledge of undergraduate EFL students' academic writing practice by exploring the semantic patterns of their RVs employment with reference to Hunston et al. (1996)'s four RV semantic categories (i.e. *Argue, Show, Find, Think*). In particular, this study intends to probe into L2 students' reporting practices through undergraduate English majors' thesis writing in the discipline of applied linguistics. The research questions for this study then are:

RQ1: Among four semantic categories of RVs, what types of RVs are preferred in Chinese undergraduate English majors' theses compared with English L1 students' academic writing?

RQ2: What and how RVs are frequently used in undergraduate L2 English majors' theses compared with L1 student academic writing?

RQ3: What pedagogical suggestions can be proposed based on the comparative analysis of RVs between L2 undergraduate students and L1 English students?

II. METHODOLOGY

A. Corpus and Text Selection

To answer the above questions, one main corpus (Undergraduate Thesis Writing Corpus, UTW) and a reference corpus (BAWE2) were compiled and processed via Sketch Engine, with comparable data size.

The UTW consists of 40 final draft theses written by the senior undergraduate English majors from 2015 to 2020, and each was revised at least three times before the final submission to ensure the maximum readability and fluency. All texts were cleaned and stored in an electronic form, and student names and any personally identifiable information were all removed. All students' first language is Chinese. To ensure generalizability, students' levels of English were not controlled. Each draft elaborated on a research topic either on Translation (*n=34*) or Culture (*n=6*), and contained introduction, literature review, theoretical framework, analysis, and conclusion.

The reference corpus (BAWE2) of English native speakers was compiled through the British Academic Written English Corpus (BAWE), a British Academic corpus of university-level student writing in the UK. It comprises texts produced with annotations of linguistics discipline and written by students with English as their first language. It can be accessed through Sketch Engine. Selecting a native student academic corpus instead of expert writing is to ensure the comparability of writing practices across groups of students. The reference corpus does not limit the levels of studies for another concern on broad comparability. Details of the two corpora can be found in Table 1.

TABLE 1
L2 LEARNERS AND L1 STUDENTS CORPUS INFORMATION

Name	Tokens	Words
UTW	310,382	~258,061
BAWE₂ (Ling+EngL1)	219,095	~183,136

B. Variables

Reporting Verbs. The first stage of the analysis involved compiling a master list of RVs for the subsequent analysis. To efficiently identify the RVs in students' theses, the master list of this study was taken from the list proposed by Kwon et al. (2018), detailing 53 RVs. In the subsequent concordance search, four Argue verbs (*assert, hypothesize, remark, posit*) were removed in that no result was returned in UTW and due to low frequency in BAWE₂. A following manual examination of each concordance line further revealed that one high-frequency RV (*add*) did not denote citation practices but rather was in relation to a translation technique, called addition or adding. This verb was also removed from the master list. All together a total of 48 RVs were searched and the correspondent concordance lines were retrieved.

Semantic Category. To examine the patterns of RVs, the framework of semantic categories was adapted from Kwon et al. (2018), which was adopted from Charles (2006b) and Friginal (2013). Applying the framework, the study retrieved the data for all 48 verb lemmas (e.g. *show, shows, showed, shown, showing*) in the four semantic categories (Argue, Show, Find, Think) in two corpora. The list of RVs by semantic category in this study is presented in Table 2. The rhetorical functions of RVs will also be examined following the study by Charles (2006a) and Charles (2006b).

TABLE 2
REPORTING VERBS BY SEMANTIC CATEGORY

Argue (27)	argue, suggest, predict, write, explain, conclude, mention, admit, observe, accept, claim, imply, complain, point out, say, insist, maintain, propose, reply, speculate, stress, contend, state, report, postulate, acknowledge, talk about
Show (7)	show, illustrate, indicate, demonstrate, confirm, mean, reveal
Find (9)	find, realize, observe, discover, establish, infer, recognize, identify, note
Think (6)	think, hold, assume, feel, hope, know

C. Research Procedure

A mixed-method comparative analysis combining both quantitative and qualitative procedures was conducted to explore L2 learners' academic reporting practices and to find any potential issues in comparison with L1 students' academic writing practices, following the procedure as follows:

(1). The comparative analysis was first made based on the overall frequency of each type of RVs across four semantic categories in each corpus. Each reporting verb was searched and frequency data was retrieved one by one based on the master list in the main corpus (UTW) and the reference corpus (BAWE2) (RQ1). Cross-references to the findings in other associated studies will be discussed to draw more insights on the RV practices for L2 students.

(2). The subsequent study summarized Top 10 frequently used RVs and manually examined the returned concordance lines to ensure the reliability of the results (RQ2). Scrutinization then zoomed in onto the concordance lines of two representative RVs in two corpora.

(4) The study then correspondingly provided pedagogical suggestions based on the findings (RQ3).

III. FINDINGS AND DISCUSSION

The session will first illustrate the overall frequency of RVs across four semantic categories, and then discuss the most frequently used individual verbs in relation to these categories. To address a confusion emerging from the inquiry, the study conducted a collocational analysis for two node words. Cross-references to the findings in Kwon et al. (2018) and Friginal (2013) will be discussed below as they used the same semantic framework as did this study. The frequency data were normalized to instances per 1000 words due to the size of each corpus for clearer presentation. In the end, pedagogical suggestions will be provided based on the comparative analysis.

A. Semantic Category of RV Employment in L2 and L1 Academic Writing

Following the procedure stated above, the frequency data of RVs in four semantic categories was retrieved and the result can be found in Table 3. Since the two corpora displayed disproportionate size of data, log likelihood was run to compare the frequency distribution across two corpora using a log-likelihood calculator[1](Rayson & Garside, 2000).

TABLE 3
FREQUENCY RATES OF REPORTING VERB ACROSS SEMANTIC CATEGORIES IN TWO CORPORA

| | UTW | | | BAWE₂ | | | | |
	F*	NF*	Percentage to all	F	NF	Percentage to all	Log-likelihood	Sig.
Argue#	1025	330.6	36.5%	1497	683.3	44.2%	329.36	0.000 *** -
Find#	434	139.8	15.5%	675	308.1	19.9%	170.20	0.000 *** -
Show#	739	238.1	26.3%	794	362.4	23.4%	67.39	0.000 *** -
Think	609	196.2	21.7%	423	193.1	12.5%	0.07	0.799 +
Total	2807	904.7	100.0%	3389	1546.8	100.0%	444.43	0.000 *** -

(*F: Raw frequency; NF: Normalized frequency)

In-group and cross-group comparison for frequency rates revealed some interesting phenomena. For starters, log-likelihood results suggested that there was **a significantly less use of RVs by L2 students than L1 students in general,** which can be particularly illustrated by three categories of RVs (i.e. Argue, Find and Show) in L2 students' academic writing. As reporting practice is part of construction of writer's stance in evidence-based academic writing, the findings, though centered on student writers, coincide with relevant previous studies on professional academic stance-taking practices in L2 English publications by Chinese researchers, implying that Chinese academic writers would often employ more confident stance as opposed to customary cautious positioning taken by Anglophone authors (e.g. Chen & Zhang, 2017; Hu & Cao, 2011; Yang, 2013).

A further vertical within-group examination in UTW found that students' preference over four semantic types of RVs followed the order by Argue > Show > Think > Find, while in BAWE2, students followed the order by Argue > Show > Find > Think. In general, it indicates **an overall preference for Argue verbs in both English L2 and L1 students in Applied Linguistics**. This was similarly found in L2 students' literature review section (Kwon et al., 2018) and L1 forestry academic papers (Friginal, 2013), yet different from the RVs choice by L1 forestry students (Friginal, 2013) (see Figure 1).
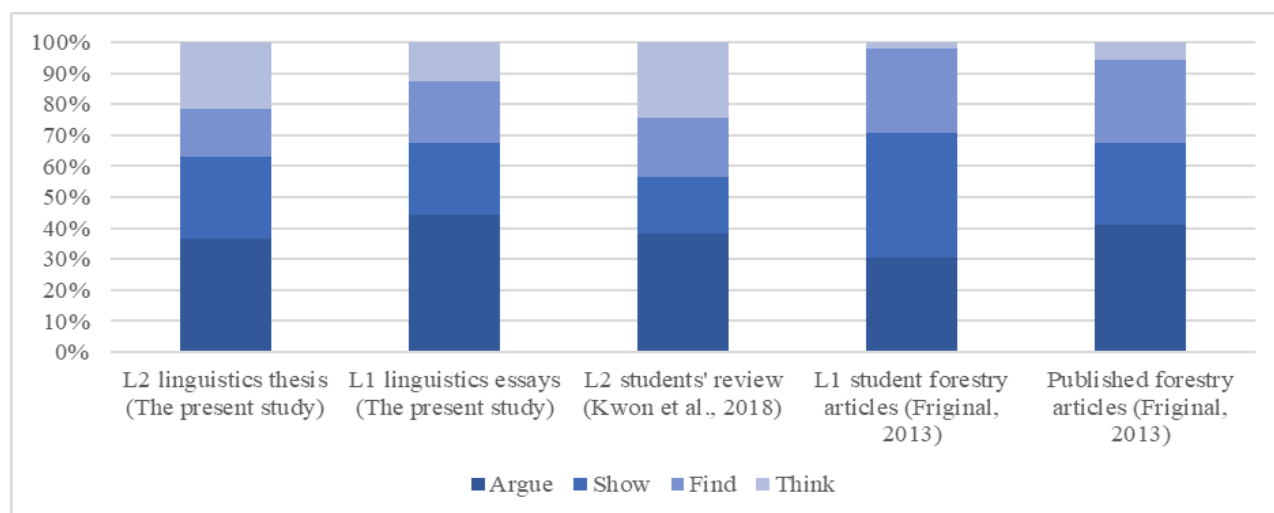

Figure 1 Cross-reference of RV Patterns in Four Semantic Categories With Kwon et al. (2018) and Friginal (2013)

Such similarity between the present study and Kwon et al. (2018) might imply that the preference for Argue verbs could be with no "disciplinary specificity" in L2 students' citation practice. However, a comparison amongst L1 linguistic essays in the present study, L1 student forestry articles (Friginal, 2013), and L1 published forestry articles (Friginal, 2013) point to **a much divergent pattern of Argue verbs at varied levels of L1 expertise**. Yet considering that the data for L1 forestry students (Friginal, 2013) was retrieved before a writing instruction session, this might suggest a highlighted need for argumentative writing training, regardless of academic components being included in such training or not. Nonetheless, it is unwarranted to generalize findings based on mere comparisons across studies without a similar baseline of contexts, even though such cross-reference might provide some insights and point to a space for further investigation.

One noticeable yet not statistically significant difference in RVs choice was found in **Think verbs** between UTW and BAWE₂ as L2 students overly relied on Think verbs (=21.7%) compared with what their L1 student peers did (=12.5%). Additionally, such cognitive reporting verbs were found to be similarly least favored in both Chinese and Anglophone academic journals (Liu & Wang, 2019). This further implied an over-reliance on intuitively persuasive

---

[1] https://ucrel.lancs.ac.uk/llwizard.html

manner amongst Chinese learners due to a **lack of rigorous training on evidence-based argumentation.** The generally less frequent use of RVs for L2 learners also echoes with such findings as insufficient training for L2 students.

B. Frequently Used RVs in L2 and L1 Academic Writing

A further analysis was conducted by summarizing the top 10 uses of RVs in both UTW and BAWE$_2$ (Table 4). A full list of frequency of all 48 RVs in two corpora can be found in Appendix. These top 10 RVs nearly account for the majority of all counts of RVs (*2/3, =64% & 60% respectively*), thus making further analysis reasonable.

TABLE 4
THE 10 MOST FREQUENT REPORTING VERBS IN TWO CORPORA

UTW (top 10)					BAWE$_2$ (top 10)				
RV	**Category**	**F**	**NF**	**Percentage of all (2807)**	**RV**	**Category**	**F**	**NF**	**Percentage of all (3389)**
show	Show	294	94.7	10.50%	show	Show	373	170.2	11.00%
mean	Show	276	88.9	9.80%	suggest	Argue	352	160.7	10.40%
know	Think	238	76.7	8.50%	find	Find	338	154.3	10.00%
find	Find	210	67.7	7.50%	say	Argue	222	101.3	6.60%
say	Argue	203	65.4	7.20%	identify	Find	143	65.3	4.20%
think	Think	177	57	6.30%	mean	Show	135	61.6	4.00%
mention	Argue	107	34.5	3.80%	write	Argue	129	58.9	3.80%
write	Argue	102	32.9	3.60%	feel	Think	127	58	3.70%
explain	Argue	91	29.3	3.20%	know	Think	111	50.7	3.30%
realize	Find	90	29	3.20%	think	Think	107	48.8	3.20%
Proportion to overall no. of RVs				64%	*Proportion to overall no. of RVs*				60%

A close examination of the favorite RVs in UTW suggested **a rather restricted repertoire of colloquial verbs in L2 students' citation practice** (e.g. *know, say*) compared with that of L1 students (e.g. *suggest, identify*). Furthermore, in UTW, the top two frequent verbs, *show* and *mean,* both belong to Show category, while the Argue verbs (i.e. *say, mention, write, explain*) were used relatively less at an inconspicuous position. By comparison, though L1 students frequently used *show*, their choice of other RVs was evenly distributed among the three categories (Show, Argue, Find) and Argue verbs (*suggest, say, write*) were at a prominent position, especially the verb *suggest* (2nd most frequent). As explained above, Show verbs indicate a fact or situation, while Argue verbs are concerned with writing and other forms of communication (Charles, 2006b). **The divergent uses of Argue and Show verbs** thus can suggest that either L2 students tend to build up their argumentation more by introducing facts rather than by manner of arguing, or that their use of RVs was incompatible with the conventionally intended semantic functions of the chosen verbs.

To elucidate such confusion and to answer the second part of RQ2 (i.e. *how do RVs are frequently used by L2 students?*), the current study undertook an *ad hoc* scrutiny of the contextual information in which two node verbs were present, particularly in terms of their collocation with subjects. Two RVs were selected— *show* for being both the most favorite RV in both corpora, and *suggest* for being the second frequent item in BAWE$_2$ yet not even making its way into the list of UTW. As the purpose of answering RQ2 is to find pedagogical suggestions for EAP writing training, conducting closer examination of the two representative RVs should suffice.

Table 5 outlines the subject-object collocation of the node verbs, i.e. *show* and *suggest*. As a typical Show verb, *show* signals writers' clear acceptance of certain findings/studies/research (Charles, 2006b); it would thus be expected to be used in research- or finding-related contexts, precisely as what L1 English students presented via their collocational patterns of *show* in BAWE$_2$ (see Table 5 and example 4). However, such collocational representation was not detected in L2 writing corpus (UTW). In UTW, although students used it in a factive manner, they seemed to show **limited access to the intended rhetorical functions** of *show* (see Table 5 and example 1-3):

(1) This requires *translators to <u>show</u> more respect* to source language and approach the original text as closely as possible. (UTW)

(2) As a speech delivered at the 23rd China-ASEAN leaders' meeting and *the speech <u>shows</u> a relatively formal style*, the scales of formality can be observed. (UTW)

(3) *The table below <u>shows</u> the differences* between the two versions in the translation in a passage of Charlotte. (UTW)

(4) *Studies have <u>shown</u>* the importance of social attitudes and class in the construction of 'prestigious' language features. (BAWE$_2$)

Moreover, the use of *suggest* by L1 and L2 students respectively demonstrated a notably divergent pattern. The Argue category of RVs is part of linguistic hedging devices, offering tentative explanations or bringing out claims in conflict or with uncertainty (Charles, 2006a, 2006b). It is thus normally accompanied by inclusion of the positions of other researchers, allowing writers to situate their own study within the perceivable field (Charles, 2006b). Concordance search for the collocation of the lemma *suggest* confirmed the use of such rhetorical function (see example 8, as *suggest* was used to implicitly situate the writer's stance between the cited studies). Table 5 also illustrated other research-related subjects being collocated with *suggest,* similar to the use in the example 8. In UTW, however, the intended rhetorical functions of *suggest* were found to be **misplaced into the rhetorical contexts** as with Show or Think verbs. For L2 students, a prominent type of collocated subjects with *suggest* directs to human agents, in particular, the writer

him/herself (see Table 5 and example 5-6). Additionally, non-human agents were not introduced for the purpose of tentative explanation, but rather explicit presentation of factual information (see example 7). There were no plausible subsequent opinions proposed through the meaning intention of tentativeness as frequently marked by *suggest* in L1 writing corpus:

(5) With the development of film industry, more subtitle groups have come into being, yet the quality of translation of different subtitle groups has been patchy. Standards and rules thus should be set up. And *the author suggests to* include academic institutions to offer courses on subtitle translation to train more qualified subtitlers. (UTW)

(6) In the writer's point of view, the expressive function also conveys information of films as film titles reveal the intentions of the directors and set the subject tone for audiences to be emotionally inspired. Therefore, *the writer suggests that* the expressive function related to the informative function. (UTW)

(7) The coherence principle, as *the name suggests*, means that the translation must conform to the expression habits of the target language. (UTW)

(8) Cappa et.al (2000) claim that neuroimaging work has shown 'right brain involvement in both language comprehension and production' (p.28). *These findings suggest that* language functions are located across the brain and may not just be limited to the left hemisphere. Having said this it is important to note that it is generally agreed that complex grammar skills are localized to only one hemisphere (Lustepal, 1995). (BAWE$_2$)

Thus, the confusion aforementioned is tentatively clarified — L2 students may not be unaware of the conventional manner of argumentation in EAP thesis writing regardless of **their misuse of RVs in normally unintended contexts.** This is indicative of a possible lack of argumentative writing training on **specific rhetorical functions of each type of reporting verbs**.

TABLE 5
SUBJECT-OBJECT COLLOCATION OF SHOW AND SUGGEST IN UTW AND BAWE2

UTW						BAWE$_2$					
Show											
Subject	*F>3*	*logDice*	*Object*	*F>3*	*logDice*	**Subject**	*F>3*	*logDice*	*Object*	*F>3*	*logDice*
example	8	10.6	respect	7	10.4	study	11	8.5	adjective	9	10.3
word	4	9.1	style	6	9.8	research	9	8.8	agreement	5	8.7
translation	4	7.2	difference	5	9.1	example	6	9.2	difference	5	7.4
version	3	9.3	reader	5	9.0	result	6	8.1	use	5	7.3
			meaning	5	8.4	experiment	5	8.5	preference	4	8.8
			connotation	4	9.2	finding	4	8.9	variation	4	8.4
			image	4	9.0	word	3	8.2	pattern	4	7.6
			formality	3	9.2	blue	3	9.3	evidence	4	7
			trend	3	9.2	red	3	9.3	way	4	6.7
			sign	3	8.9	figure	3	8.1	result	4	6.2
			feature	3	8.6	language	3	7.9	lesion	3	8.9
			culture	3	8.2	evidence	3	7.3	understanding	3	7.4
									process	3	6.1
									change	3	6.1
Suggest											
Subject	*F.>2**	*logDice*	*Object*	*F>2*	*logDice*	**Subject**	*F>2**	*logDice*	*Object*	*F>2*	*logDice*
author	4	9.1				finding	6	9.74	child	5	7.7
principle	2	10.4				research	5	8.03	cause	2	7.8
name	2	10.4				theory	5	7.56	reason	2	7.3
writer	2	10.1				sociolinguists	2	9.25	way	2	5.8
						question	2	7.47	woman	2	5.8
						evidence	2	6.81	theory	2	5.8

(*minimum frequency set as 2 due to fewer uses of "suggest" in UTW)

C. Pedagogical Implications from the Comparative Analysis

The above analysis points to a macro- to micro-level pedagogical implications that demand further attention in future academic writing training.

There is a lack of rigorous academic training noticed via the significantly lower deployment of RVs across three categories (Argue, Show, Find) and an over-dependence on instinctively persuasive manner amongst Chinese learners of L2 English. Curriculum designers of EAP or even basic-level second/foreign language writing courses should be particularly aware of the current need to promote the instruction of evidence-based argumentative (inferential or deductive) skills. As strong academic-oriented argumentative writing naturally involves a synthesis of other writers' findings and adoption of a carefully articulated stance toward cited statements, it is essential for teaching practitioners to highlight the rhetorical functions of various RV types in their academic writing instructions (also see Jarkovská & Kurková, 2021).

Furthermore, this study also observes that students demonstrate a limited lexical repertoire of conversational RVs and that their use of only a few RVs was misconstrued in specific contexts. This heightens the need for EAP instructors to equip L2 students with a variety of reporting devices as well as a better understanding of the lexical or perhaps grammatical aspects of citation (also see Jarkovská & Kurková, 2021; Nguyen & Pramoolsook, 2016). It would be

beneficial to novice writers if textbook designers could integrate rhetorical functions and authentic usage of reporting practices from native and/or expert writers, as this would allow student-writers to obtain access to academic writing in an authentic context.

IV. CONCLUSION

Based on the semantic categories of Hunston et al. (1996) and Charles (2006a), this study sets out to explore the semantic patterns of RVs in undergraduate theses written by L1 Chinese English majors. Through comparative analysis, the study found an overall preference for Argue verbs in both English L2 and L1 students. The examination of Think verbs between the two corpora suggests a seemingly intuitive argumentative manner and significantly less frequent use of RVs in three categories (Argue, Find and Show), both pointing to insufficient EAP writing training.

Further summarization of the 10 most frequent RVs in UTW and BAWE$_2$ reveals a restricted repertoire of colloquial verbs in L2 students' citation practice. A collocational analysis of two verbs (*show, suggest*) was then conducted to address a confusion emerging from the inconsistent preference for Show and Argue verbs in the two corpora. It indicates students' preliminary awareness of argumentative manner, but their use of RVs was misrepresented in context.

All of the above findings underline the central theme of the study— L2 English majors, though mostly trained to write argumentative essays, are still in need of further instructions on academic-driven argumentative writing. Pedagogical suggestions include more rigorous training on evidence-based argumentative processes, intensive language training on academic lexicons, and deepening academic writing practices with timely feedback.

The present study has limitations. Future studies could expand the master list (see Liardét & Black, 2019), and efforts shall be made to closely examine the contextual information where RVs are in use, for which Hyland (1999)'s framework can be adopted. It is also unclear what voices and tenses of RVs are used by L2 English learners. In addition, discrepancy on RV employment is not a single-factor outcome, instead it should factor multi-variables (e.g. learners and experts, L1 and L2 English users, disciplinary difference) into the final linguistic representation. Investigations on other issues related to L2 undergraduate English learners are certainly unexhausted through the analysis of the study.

Nonetheless, by revealing some findings on L2 undergraduate academic writing behaviors, the study managed to provide two-fold pedagogical suggestions for undergraduate EAP writing course designers. The framework used in this study can also be good entry points for further investigations on other related concerns and it is the author's hope that this study can offer some insights for understanding L2 undergraduate EAP writing practices.

APPENDIX. FULL LIST OF FREQUENCY OF REPORTING VERBS IN TWO CORPORA

SC*	RV	UTW		BAWE2		Log-likelihood	Sig.		
		F*	NF*	F	NF				
Show	show	294	947.22	373	1702.46	57.02	0.000	***	-
Show	mean	276	889.23	135	616.17	12.66	0.000	***	+
Think	know	238	766.8	111	506.63	13.59	0.000	***	+
Find	find	210	676.59	338	1542.71	91.29	0.000	***	-
Argue	say	203	654.03	222	1013.26	20.29	0.000	***	-
Think	think	177	570.27	107	488.37	1.62	0.203		+
Argue	mention	107	344.74	38	173.44	14.54	0.000	***	+
Argue	write	102	328.63	129	588.79	19.54	0.000	***	-
Argue	explain	91	293.19	83	378.83	2.83	0.092		-
Find	realize	90	289.97	0	0				
Think	feel	89	286.74	127	579.66	26.47	0.000	***	-
Argue	propose	71	228.75	43	196.26	0.64	0.425		+
Argue	conclude	68	219.08	64	292.11	2.71	0.100		-
Think	hold	65	209.42	39	178	0.65	0.419		+
Show	indicate	64	206.2	81	369.7	12.29	0.000	***	-
Argue	point out	61	196.53	45	205.38	0.05	0.823		-
Argue	accept	54	173.98	22	100.41	5.05	0.025	*	+
Argue	suggest	46	148.2	352	1606.61	385.35	0.000	***	-
Argue	talk about	40	128.87	27	123.23	0.03	0.857		+
Find	establish	39	125.65	25	114.11	0.14	0.706		+
Show	reveal	38	122.43	26	118.67	0.02	0.902		+
Think	hope	36	115.99	4	18.26	19.51	0.000	***	+
Argue	maintain	34	109.54	34	155.18	2.05	0.152		-
Show	illustrate	33	106.32	59	269.29	19.28	0.000	***	-
Argue	state	30	96.66	105	479.24	74.33	0.000	***	-
Show	demonstrate	28	90.21	105	479.24	78.31	0.000	***	-
Argue	stress	28	90.21	12	54.77	2.22	0.137		+
Argue	imply	24	77.32	28	127.8	3.27	0.071		-
Find	identify	22	70.88	143	652.68	146.28	0.000	***	-
Find	recognize	20	64.44	2	9.13	11.49	0.001	***	+
Find	discover	19	61.21	31	141.49	8.60	0.003	**	-
Argue	claim	18	57.99	61	278.42	42.09	0.000	***	-
Find	infer	14	45.11	10	45.64	0.00	0.977		-
Argue	insist	10	32.22	2	9.13	3.40	0.065		+
Find	note	10	32.22	73	333.19	78.44	0.000	***	-
Argue	observe	10	32.22	53	241.9	49.08	0.000	***	-
Find	observe	10	32.22	53	241.9	49.08	0.000	***	-
Argue	argue	8	25.77	42	191.7	38.70	0.000	***	-
Show	confirm	6	19.33	15	68.46	7.75	0.005	**	-
Think	assume	4	12.89	35	159.75	40.25	0.000	***	-
Argue	reply	4	12.89	2	9.13	0.16	0.685		
Argue	complain	3	9.67	4	18.26	0.70	0.402		
Argue	postulate	3	9.67	6	27.39	2.34	0.126		
Argue	acknowledge	2	9.67	23	104.98	28.79	0.000	***	-
Argue	predict	2	6.44	20	91.28	24.03	0.000	***	-
Argue	report	2	6.44	54	246.47	80.18	0.000	***	-
Argue	speculate	2	6.44	0	0				
Argue	admit	1	3.22	8	36.51	8.91	0.003	**	-
Argue	contend	1	3.22	0	0				
Argue	assert	0	0	13	59.33				
Argue	hypothesize	0	0	2	9.13				
Argue	posit	0	0	3	13.69				
Argue	remark	0	0	0	0				
Total=		2807	9046.93	3389	15468.1				

*SC: semantic category; F.: Frequency; NF: normalized frequency/1,000

ACKNOWLEDGEMENTS

REFERENCES

[1] Charles, M. (2006a). The Construction of Stance in Reporting Clauses: A Cross-disciplinary Study of Theses. *Applied Linguistics, 27*(3), 492-518. https://doi.org/10.1093/applin/aml021

[2] Charles, M. (2006b). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes, 25*(3), 310-331. https://doi.org/10.1016/j.esp.2005.05.003

[3] Chen, C., & Zhang, J. L. (2017). An intercultural analysis of the use of hedging by Chinese and Anglophone academic English writers. *Applied linguistics review, 8*(1), 1-34. https://doi.org/10.1515/applirev-2016-2009

[4] Friginal, E. (2013). Developing research report writing skills using corpora. *English for Specific Purposes, 32*(4), 208-220. https://doi.org/10.1016/j.esp.2013.06.001

[5] Granger, S., & Paquot, M. (2009). Lexical Verbs in Academic Discourse: A Corpus-driven Study of Learner Use. In M. Charles, D. Pecorari, & S. Hunston (Eds.), *Academic Writing*: At the Interface of Corpus and Discourse (1 ed., pp. 193-214). Bloomsbury Academic. https://doi.org/10.5040/9781474211703.ch-010

[6] Hinkel, E. L. I. (2003). Simplicity Without Elegance: Features of Sentences in L1 and L2 Academic Texts. *TESOL Quarterly, 37*(2), 275-301. https://doi.org/10.2307/3588505

[7] Hunston, S.& Francis, G., & Manning, E. (1996). *Collins COBUILD grammar patterns 1: Verbs*. HarperCollins.

[8] Hu, G., & Cao, F. (2011). Hedging and boosting in abstracts of applied linguistics articles: A comparative study of English- and Chinese-medium journals. *Journal of Pragmatics, 43*(11), 2795-2809. https://doi.org/10.1016/j.pragma.2011.04.007

[9] Hyland, K. (1999). Academic attribution: citation and the construction of disciplinary knowledge. *Applied Linguistics, 20*(3), 341-367. https://doi.org/10.1093/applin/20.3.341

[10] Hyland, K., & Jiang, F. (2019). Points of Reference: Changing Patterns of Academic Citation. *Applied Linguistics, 40*(1), 64-85. https://doi.org/10.1093/applin/amx012

[11] Jarkovská, M., & Kučírková, L. (2021). Reporting Verbs and Related Syntactic Choices in Students' Theses: A Study of Two Disciplines. *Journal on Efficiency and Responsibility in Education and Science, 14*(3), 130-142. https://doi.org/10.7160/eriesj.2021.140301

[12] Kwon, M. H., Staples, S., & Partridge, R. S. (2018). Source work in the first-year L2 writing classroom: Undergraduate L2 writers' use of reporting verbs. *Journal of English for Academic Purposes, 34*, 86-96. https://doi.org/10.1016/j.jeap.2018.04.001

[13] Liardét, C. L., & Black, S. (2019). "So and so" says, states and argues: A corpus-assisted engagement analysis of reporting verbs. *Journal of Second Language Writing, 44*, 37-50. https://doi.org/10.1016/j.jslw.2019.02.001

[14] Liu, R., & Wang, S. (2019). *How Citation Is Signaled*: A Corpus-Based Study on Reporting Verbs in Chinese Academic Papers. In (pp. 53-71). Singapore: Springer Singapore. https://doi.org/10.1007/978-981-13-9505-53

[15] Marti, L., Yilmaz, S., & Bayyurt, Y. (2019). Reporting research in applied linguistics: The role of nativeness and expertise. *Journal of English for Academic Purposes, 40*, 98-114. https://doi.org/10.1016/j.jeap.2019.05.005

[16] Miller, R. T., Mitchell, T. D., & Pessoa, S. (2014). Valued voices: Students' use of Engagement in argumentative history writing. *Linguistics and Education, 28*, 107-120. https://doi.org/10.1016/j.linged.2014.10.002

[17] Nguyen, T. T. L., & Pramoolsook, I. (2016). Citations in literature review chapters of TESOL master's theses by Vietnamese postgraduates, *GEMA Online Journal of Language Studies, 16* (2), 17–32. https://doi.org/10.17576/gema-2016-1602-02

[18] Ramoroka, B. T. (2014). Integration of sources in academic writing: A corpus-based study of citation practices in essay writing in two departments at the University of Botswana. *Reading & Writing, 5*(1). https://doi.org/10.4102/rw.v5i1.41

[19] Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. *The workshop on comparing corpora*, 1-6

[20] Shaw, P. (1992). Reasons for the Correlation of Voice, Tense, and Sentence Function in Reporting Verbs. *Applied Linguistics, 13*(3), 302-319. https://doi.org/10.1093/applin/13.3.302

[21] Thompson, G., & Ye, Y. (1991). Evaluation in the Reporting Verbs Used in Academic Papers. *Applied Linguistics, 12*(4), 365–382. https://doi.org/10.1093/applin/12.4.365

[22] Uba, S. Y. (2019). Semantic Categories of Reporting Verbs across Four Disciplines in Research Articles. *English Language Teaching, 13*(1). https://doi.org/10.5539/elt.v13n1p89

[23] Yang, Y. (2013). Exploring linguistic and cultural variations in the use of hedges in English and Chinese scientific discourse. *Journal of Pragmatics, 50*(1), 23-36. https://doi.org/10.1016/j.pragma.2013.01.008

[24] Yeganeh, M. T., & Boghayeri, M. (2015). The Frequency and Function of Reporting Verbs in Research Articles Written by Native Persian and English Speakers. *Procedia - Social and Behavioral Sciences, 192*, 582-586. https://doi.org/10.1016/j.sbspro.2015.06.097

Yueyue Huang (BA, MA) is currently a doctorate student at the Hong Kong Polytechnic University (Hong Kong, China) in the field of Applied Language Sciences, and a lecturer at Guangzhou Xinhua University (Guangdong, China). Her research interests include corpus-based translation and interpreting Studies as well as second language reading and writing.