

A Corpus-Based Comparative Study of Exemplification in Hard and Soft Sciences Research Articles: A Case of Chinese and English Authors

Yang Yuanyuan ^a

yangyuanyuan@graduate.utm.my

Language Academy,

Universiti Teknologi Malaysia, Malaysia

&

School of Foreign Languages,

Shangqiu Normal University, China

Noor Mala Ibrahim ^b

m-nmala@utm.my

Language Academy,

Universiti Teknologi Malaysia, Malaysia

ABSTRACT

Exemplification, an essential part of argumentation in academic writing, helps authors to support general ideas with specific examples. This corpus-based investigation is a preliminary comparative study on the use of exemplification by Chinese and English native authors in research articles (RAs) of hard sciences and soft sciences covering six different disciplines. The study focuses primarily on three aspects: Exemplifying Markers (EMs), exemplifying units, and functions of exemplification. A corpus of twelve RAs with a total number of 143,604 words across six disciplines was built with six articles each from the hard sciences and the soft sciences. Triki's (2021) model of exemplification was used as the analytical framework. The exemplification theory serves as the theoretical basis for the study and K-means clustering and Welch Two Sample t-tests are employed as the research methods. Findings show that diverse categories of EMs are used by authors from these two different linguistic backgrounds and in different disciplines. It is also revealed that the authors from the two linguistic backgrounds use EMs more similarly in hard sciences than in soft sciences, even though EMs are usually used more frequently in the soft sciences. In addition, there are significant differences in the use of EMs between the two groups of authors and different disciplines. More exemplifying groups than exemplifying clauses are employed across linguistic backgrounds and disciplines. Various categories of functions are performed by exemplification. The findings may be valuable in guiding exemplification instruction and learning in English for Academic Purposes (EAP) environments.

Keywords: exemplification; corpus; research article; linguistic background; hard and soft sciences

INTRODUCTION

Publication of research articles (RAs) in international journals has grown in importance as a criterion for assessing a scholar's academic achievement. The internationalization tendency, thus the obligation to publish in English (Crystal, 2003), and the regional norms and values, i.e. the "local language codes, cultures, and ideologies of literacy", underlying authors' linguistic

^a Main author

^b Corresponding author

background for publication, both have a significant impact on the publishing activity (Yuan, 2021, p.5). The paper publication in English-language journals has therefore become the main means of disseminating and advancing scientific research. As a result, being able to write well in English is one of the essential qualifications for participating in international academic communication, in the dissemination of scientific knowledge.

Within the genre of RAs, there are certain communicative objectives to be met (Jasim, 2023) and Hyland (2005) has emphasized the importance of the writing style of an author for his or her RA to be accepted for publication. Exemplification serves as one of the crucial rhetorical devices that aid in the clarification, elaboration, and support of arguments in RAs. According to Hyland (2007), exemplification is a communication procedure in which meaning is clarified or strengthened by a second unit that exemplifies the first through the use of an example. Exemplification also shows writers' beliefs about readers' needs and how they will respond to them, hence it is critical to the reader-text interaction.

The use of exemplification, however, can vary across languages and disciplines, potentially impacting the clarity and effectiveness of scholarly communication. Thus, exemplification merits an in-depth study. Siepmann (2005, p.112) names three particularly notable aspects. The first pertains to exemplification as a complex discourse technique presented virtually in every argumentative text. The second aspect relates to its frequent appearance in academic texts. The third concerns the challenges it poses for both L1 and L2 learners. As Siepmann highlights, "writers at an advanced stage of language learning [...] experience no difficulty in forming and using exemplifiers, the evidence, however, clearly suggests otherwise" (Siepmann, 2005, p.257). Paquot's (2008) study has also demonstrated that English L2 writers have a far smaller vocabulary reserved of exemplifying phraseologies than their English L1 counterparts. Clearly, exemplification is not a simple discourse function but a complex rhetorical category (Hyland, 2007).

This paper aims to compare the patterns of exemplification employed by Chinese and native English speakers in RAs within the realms of soft and hard sciences. By comparing the usage of exemplification in these two linguistic and disciplinary contexts, we seek to shed light on variations of EMs used, their classification based on semantic and syntactic categories and finally the functions that they serve. The application of exemplifying units is also investigated. Conducting a thorough analysis of exemplification in academic writing is an important task in and of itself, as it could contribute to the systematic documentation and description of the lexicogrammatical resources that can be used for exemplification. The findings of these studies may also be useful in guiding exemplification instruction and learning in EAP environments, notably in China.

PREVIOUS STUDIES ON EXEMPLIFICATION

Exemplification is a communication technique used by authors to make their ideas more accessible by providing explanation, illustration, or reinforcement. It is the process of explaining or clarifying a claim (which could be a theory, an observation, or an argument) or providing sufficient details to prove a superordinate category by, for example, displaying subordinate categories (cf. Hyland, 2007; Paquot, 2008). Exemplification is an essential part of exposition and academic writing as it is found in almost any text of argumentation (Siepmann, 2005). A writer (or a speaker) conducts the act of giving an example by discursively bringing up a particular item, person, or circumstance as an illustration of a generalized idea or principle in a communication process known as

elaboration. The “translatability between generality and particularity” is the attribute of such exemplifying activities (Harvey, 2002, p.viii).

Previous research on exemplification have covered various areas, particularly news reporting (Bigsby et al., 2019; Kramer & Peter, 2020; Zillmann et al., 1996) as well as exemplification in dictionaries (Liu, 2017; Xu, 2008; Hu, 2001) and exemplification in languages other than English, such as Japanese (Barotto, 2021; Taylor, 2010), Danish (Petersen, 2020) and Chinese (Su & Fu, 2023). In the academic discipline, while exemplification has been considered an ubiquitous feature in academic writing (Su & Zhang, 2020), Triki (2021) claims that there has been little attention given to the use of exemplification in academic writing. Su et al. (2022) agree that exemplification in EAP has been a neglected area. Many studies on exemplification in the academic field have been in relation to the study on metadiscourse (e.g. Hyland, 2005; Barotto, 2018a; Guziurová, 2022) as exemplification is one of the two categories of Code Glosses in Hyland’s framework of metadiscourse.

Among the investigations on exemplification in academic disciplines, various studies have looked at exemplification in students’ writing. For example, Hinkel (2001) looked at the use of examples and illustrations in academic essays produced by students from various language backgrounds such as English, Chinese, Japanese, Korean, Indonesian and Vietnamese during a diagnostic test for a degree program in four American universities. He discovered that the non-native students used twice as many exemplifications as their native counterparts. In a different study, Paquot (2008) investigated the use of exemplification among native students using Louvain Corpus of Native Speaker Essays (LOCNESS) (Granger, 1996) and non-native students using five sub-corpora of the International Corpus of Learner English (ICLE). It was found that the non-native students overused the markers *for example* and *for instance*, suggesting the limited repertoires that non-native students have when introducing arguments and points of view. The learners ‘cling on’ to certain fixed phrases and expressions which they feel confident in using (Granger, 1996).

Interestingly, Guziurová (2022) investigated the use of exemplifications in the theses among Czech L1 postgraduate students in the fields of linguistics, literature and English Language Teaching (ELT) Methodology, and found that the postgraduate students heavily used *i.e.* and *such as*. She concluded that even though the postgraduate students recognized the importance of exemplification and made use of many exemplifications in their writing, they relied on simple grammaticalized forms that do not require much processing efforts. In a similar vein, Su et al. (2022) conducted a study comparing the use of exemplification by Chinese English-major MA students and expert writers from the local grammar perspective. They also found that the postgraduate MA students tended to rely on more frequent or typical patterns of exemplification and concluded that this preference could stem from the students’ limited repertoire of exemplification phraseology or their lesser familiarity with employing less common or atypical exemplification patterns. The findings of these studies imply the importance of raising awareness and providing support for students to develop a diverse range of exemplification strategies, enhancing the effectiveness and variety of their academic writing.

Studies specifically investigating the use of exemplification in RAs have majorly focused on the domain of soft sciences, particularly Linguistics and Applied Linguistics, and dominated by the work of Hang Su and associates. For example, Su et al., (2021) conducted a diachronic investigation on exemplification in Linguistics RAs which could “enrich the description of academic discourse and would also have useful implications for EAP and English for Specific Purposes (ESP) pedagogy” (p. 120). Then, Su et al. (2022) also compared the use of

exemplification in RAs and theses in Applied Linguistics. Other related studies include Su and Zhang (2020), Su et al. (2021) and Su and Lu (2022). As mentioned earlier, exemplification has also been investigated alongside reformulation as two sub-categories of Code Glosses. However, to the best of our review, the areas remain similar. Rahimpour (2013) studied how English and Persian academic writers used exemplification in Applied Linguistic RAs. Dehghan and Chalak (2016) examined how Iranians and native English speakers used exemplifications in the introduction section of Applied Linguistics articles, which they found no significant difference in terms of frequency of exemplifications used. Kafes (2022) also examined code glosses in a corpus of 68 Applied Linguistics RAs exploring how expert and novice writers elaborated ideas to address their readers' needs. Safari (2018) did a slightly different approach when he compared exemplification and reformulation in Political Science and Applied Linguistics RAs.

As academic disciplines are commonly placed under the umbrella terms 'hard' and 'soft' (Nesi & Gardner, 2006; Hyland, 2015), Triki (2021) advocates the analysis of exemplification across the continuum of hard and soft sciences to further understand how exemplification is manifested in academic domain. She conducted a research examining four disciplines each in hard sciences and soft sciences, and found that exemplifications are much more commonly used in soft sciences than in hard sciences. Overall, *e.g.* ranks first, followed by *such as* and *for example*. As Triki (2021) claims that native and non-native distinction was not considered while randomly selected the peer-reviewed articles that form her corpus, not much can be commented on the conclusions of other research (e.g. Paquot, 2008; Guziurova, 2022) which suggest that the use of *such as* and *for example* in non-native writing was a result of limited repertoire of exemplification among the latter. Nonetheless, all these studies have highlighted the prominent features of exemplification in academic writing and more studies should be conducted to examine this useful pedagogical resource in academic writing. This paper aims to contribute to the knowledge of exemplification in RAs in two domains of academic realms with these objectives:

1. To investigate the EMs used by authors from various linguistic backgrounds and disciplines,
2. To compare and analyze exemplifying units in the research articles from various linguistic backgrounds and disciplines, and
3. To examine the various functions of exemplification in the research articles.

METHODOLOGY

The section introduces the methodology of the current research, covering the research design, theoretical framework, the corpus, and data analysis.

RESEARCH DESIGN

To investigate the use of exemplification in research articles from various linguistic backgrounds and disciplines, a corpus-based mixed method research approach which incorporates aspects of quantitative and qualitative research is applied. For quantitative research, softwares Abbyy Finereader (<https://pdf.abbyy.com/>) and PowerGREP (<https://www.powergrep.com/>) are used to process the text and retrieve data. The articles in PDF format are recognized with the software Abbyy Finereader and turned into a corpus of TXT Documents. PowerGREP is a tool on Windows platform. Its development purpose is to help computer programmers search information in text files or coding files. PowerGREP has Search, Replace, File Finder, Collect, Sequence and other

functions. Search, Replace and Collect functions are very useful for corpus construction and corpus information extraction. Other features of PowerGREP are that it can operate on multiple directories and support the Regular Expression function. The software is valid, reliable and very powerful for corpus processing and foreign language learning. With the support of Regular Expression function, the corpus is indexed by PowerGREP and the data can be retrieved based on Triki's (2021) 25-list of EMs.

In addition, R is an open-source programming language and environment that is widely used for statistical computing and graphics (<https://cran.r-project.org/bin/windows/base/>). For the analysis of this research, K-means clustering analysis and Welch Two Sample t-tests are conducted with R language, and all the figures are drawn with R language. For the qualitative research, using Hyland's (2007) model, types of functions of exemplification are manually analyzed. Due to the integration of the advantages of both methods, mixed methods can help us obtain a more comprehensive picture than a solitary quantitative or qualitative investigation.

THEORETICAL FRAMEWORK

The theoretical basis of the current study derives from the Exemplification theory, which concerns with how incomplete depictions of more complex phenomena or populations affect subsequent social representations and activities (Zillmann, 2002; Zillmann & Brosius, 2012). It is assumed in the exemplification theory that the "compression, storage, and retrieval of elemental, concrete events are generally superior to those of complex abstract events" (Zillmann, 2002, p.25). It is also assumed in the theory that "events of consequence attract more attention and are more vigorously processed than irrelevant events" (Zillmann, 2002, p.26). The more available examples are, the more vivid and salient the information is. In other words, it can be inferred that, the application of exemplification may influence readers' perceptions and evaluations on the topic. Thus, examples can create a compelling and trustworthy piece of study.

In academic communication, particularly of research articles, theories and outcomes remain ambiguous without examples. Therefore, authors employ examples to reduce readers' uncertainties and satisfy their expectations. Examples used by the authors can illustrate the possible case and preclude the use of a counterexample. They adhere to their disciplinary norms and are understandable to readers, thus ensuring illustrations which are revealing, representative and effective.

THE CORPUS AND DATA ANALYSIS

For the corpus, RAs written in English by authors from two linguistic backgrounds, namely Chinese and English native speakers, were downloaded from the top journals in the hard and soft sciences disciplines from the Web of Science database in the period from 2021 to 2023. RAs from each of three disciplines of hard sciences – computer science, electronics, and chemistry – authored solely by Chinese writers and solely by English native writers are chosen. Similarly, RAs from three disciplines of the soft sciences – applied linguistics, education, and psychology – written solely by Chinese authors and English native authors are selected. The sampling of the RAs for the corpus is purposive and the selection of the papers is according to regularities. For the purpose of this study, only one article per discipline per linguistic background is randomly chosen. The ranking of the journals from hard sciences, soft sciences and sub-disciplines are carefully considered. The ranking is based on the relatively high Impact Factors of the journals: *ACM*

Transactions on Database Systems (Impact Factor 1.629), *IEEE Transactions on Industrial Electronics* (Impact Factor 8.162), *Nature Materials* (Impact Factor 47.656), *Applied Linguistics* (Impact Factor 3.6), *Review of Educational Research* (Impact Factor 9.24), *Psychological Review* (Impact Factor 8.247). Next, the nationalities of the authors are taken into account according to the author description from the database. Table 1 below summarizes the corpus of the study. Please note that the RAs written by the Chinese writers has an initial Cn for easier reference. In the table, the disciplines and names of the selected journals are displayed, the number of words from each journal, the number of words from each sub-corpus and the total number of words from the whole corpus are listed.

TABLE 1. The corpus of research articles

	Disciplines / Journal	Number of words	Total number of words
Hard Native	Computer/ACM Transactions on Database Systems	25298	36707
	Electronics/IEEE Transactions on Industrial Electronics	5452	
	Chemistry/Nature Materials	5957	
Soft Native	Applied Linguistics/Applied Linguistics	11972	36458
	Education/Review of Educational Research	11827	
	Psychology / Psychological Review	12659	
Hard Chinese	Computer/ACM Transactions on Database Systems	17241	27088
	Electronics/IEEE Transactions on Industrial Electronics	4918	
	Chemistry/Nature Materials	4929	
Soft Chinese	Applied Linguistics/Applied Linguistics	7517	43351
	Education/Review of Educational Research	11765	
	Psychology / Psychological Review	24069	
Total		143604	143604

(Cn is the abbreviation of Chinese)

The frequency of EMs in various disciplines used by Chinese and English native speakers was then counted, and a K-means clustering analysis of the EMs was performed to discover the connections among them. K-means clustering is a method used to group data elements based on particular similarities, resulting in the formation of clusters. The primary objective of the K-means

clustering is to group data points with similarities, which enables the identification of underlying patterns.

Next, Welch Two Sample t-tests were conducted to determine whether there are any differences in the use of EMs, by Chinese authors in hard science RAs versus in soft science RAs, as well as the use of EMs by the English native authors in hard science and soft science RAs. The Welch Two Sample t-test is a statistical method employed to determine whether the means of two populations are equal. It is typically applied when there is a noticeable difference in the variations of two populations.

The identified EMs were then divided into various semantic and syntactic categories, leading to a classification of the EMs. According to the degree of emphasis an Exemplifying Marker (EM) adds to the example it presents, EMs can be classified into four different groups from a semantic perspective. In addition to the semantic categories, three syntactic categories of exemplification are presented according to the position of the EM in comparison to that of the Exemplifying Element (EE). Then, to have a thorough understanding of exemplification, the exemplifying units, i.e., the exemplifying groups and exemplifying clauses, used in various disciplines were analyzed. Finally, an analysis of the functions that exemplification served in RAs was conducted. The functions are based on Hyland's (2007) classification, which includes (1) offering an instance to a general category; (2) providing a parallel or similar case; (3) giving a precept or a rule.

FINDINGS AND DISCUSSION

Exemplification used in hard and soft sciences RAs by authors from two different linguistic backgrounds is looked into in this study. In this section, the findings are discussed in details.

FREQUENCIES OF EMS

Among the 25 EMs based on Triki's (2021) model of exemplification, four markers, *a case in point*, *a few studies*, *a certain study*, and *to cite/mention a few*, do not appear in the corpus. Tables 2 and 3 show the frequency of the other 21 markers in English native author articles and Chinese author articles in various disciplines normalized to 10,000 words, respectively.

TABLE 2. Frequencies of EMs in English native author articles
 (normalized to 10,000 words)

Exemplifying Markers	Computer	Electronics	Chemistry	Applied Linguistics	Education	Psychology	Sum
another	4	2	0	2	3	4	15
as in	0	0	0	1	0	5	6
e.g./eg	0	0	0	2	26	4	32
example of	0	0	0	0	0	1	1
exemplify	0	0	2	0	0	0	2
extract	0	0	0	0	1	1	2
for example	2	0	5	19	27	1	54
for instance	2	0	3	1	1	2	9
in particular	0	0	0	1	2	1	4
includ*	11	7	5	9	46	7	85

like	0	0	0	0	3	3	6
mainly	0	0	0	0	0	0	0
namely	0	0	0	0	1	4	5
one such	0	0	0	0	0	0	0
particularly	0	0	0	2	0	0	2
sample	2	0	10	1	0	1	14
several/some studies	0	0	0	0	1	0	1
say	0	0	0	0	0	0	0
specifically	0	0	0	4	15	3	22
such as	15	2	7	18	19	3	64
illustrat*	2	13	3	3	9	1	31
TOTAL	38	24	35	63	154	41	355

Based on both Tables 2 and 3, a number of notable findings could be identified. First, the use of exemplifications is more prevalent among the Chinese authors comparatively with their English native counterparts. Next, a clear contrast is obvious when looking at the frequency of exemplification markers across academic domains. Specifically, higher occurrences of exemplifications are found in the soft science RAs than in the hard science for both groups of authors. When looking at exemplification within specific disciplines, it is evident that authors from both linguistic backgrounds tend to employ exemplification the most in the Education field, followed by the Applied Linguistic discipline and Psychology for the English native authors. As for the Chinese authors, the use of exemplification is exhibited more in Psychology than in Applied Linguistics. In the hard sciences, the patterns of use of exemplifications among authors of different linguistic backgrounds have been markedly different. While the English native authors use many more exemplifications in the computer field, the Chinese authors show a total opposition – exemplification in the field of computer is the least. The Chinese authors strikingly exhibit the use of exemplification the most in the field of electronics, which is interestingly a contrast to the least use among the English native authors.

In terms of individual markers, the findings have revealed a striking similarity in the use of the most prevalent types of markers. Both English native authors and Chinese authors exhibit a strong preference to use the lemma *includ** when doing exemplification. It can be seen from both Table 2 and Table 3 that the lemma *includ** has the highest frequency for both groups of authors (85 and 86 respectively). For the Chinese authors, the use of *e.g.* is as high as the lemma *includ**, while *such as* (47) and *extract* (37) rank the next two most commonly used EMs. As for the English native authors, *such as* is the second most commonly employed exemplification marker with a frequency count of 64, and *for example* is the third (54). The use of *e.g.* ranks the fourth (32). What adds intrigue to this finding is that for both groups of authors, the top four markers are notably about the same, with variations in terms of their frequencies. It is also evident that the top four markers used by each group of authors are more frequent within the realm of soft sciences rather than hard sciences.

TABLE 3. Frequencies of EMs in Chinese author articles
 (normalized to 10,000 words)

Exemplifying Markers	Cn Computer	Cn Electronics	Cn Chemistry	Cn Applied Linguistics	Cn Education	Cn Psychology	Sum
another	4	0	6	5	5	4	24
as in	0	0	0	0	4	0	4
e.g./eg	3	0	0	0	42	41	86
example of	0	0	0	0	0	0	0
exemplify	0	0	0	0	0	0	0
extract	0	0	4	1	18	14	37
for example	1	2	10	5	4	11	33
for instance	2	0	0	9	3	0	14
in particular	5	0	2	0	3	0	10
includ*	7	12	2	25	30	10	86
like	1	4	0	1	0	2	8
mainly	0	8	2	4	0	6	20
namely	1	0	0	0	1	3	5
one such	0	0	0	0	1	0	1
particularly	0	0	0	0	0	0	0
sample	1	0	2	1	0	1	5
several/some studies	0	0	0	0	3	0	3
say	3	0	0	3	0	1	7
specifically	1	0	0	0	8	9	18
such as	0	6	12	16	8	5	47
illustrat*	2	13	3	3	9	1	31
TOTAL	29	45	43	73	139	108	437

K-MEANS CLUSTERING OF EMS

The EMs used in the six disciplines by English native speakers and Chinese authors are then grouped using the K-means clustering approach in accordance with the aforementioned data. As mentioned, K-means clustering can reveal the underlying patterns of the data points. Figure 1 illustrates the outcome.

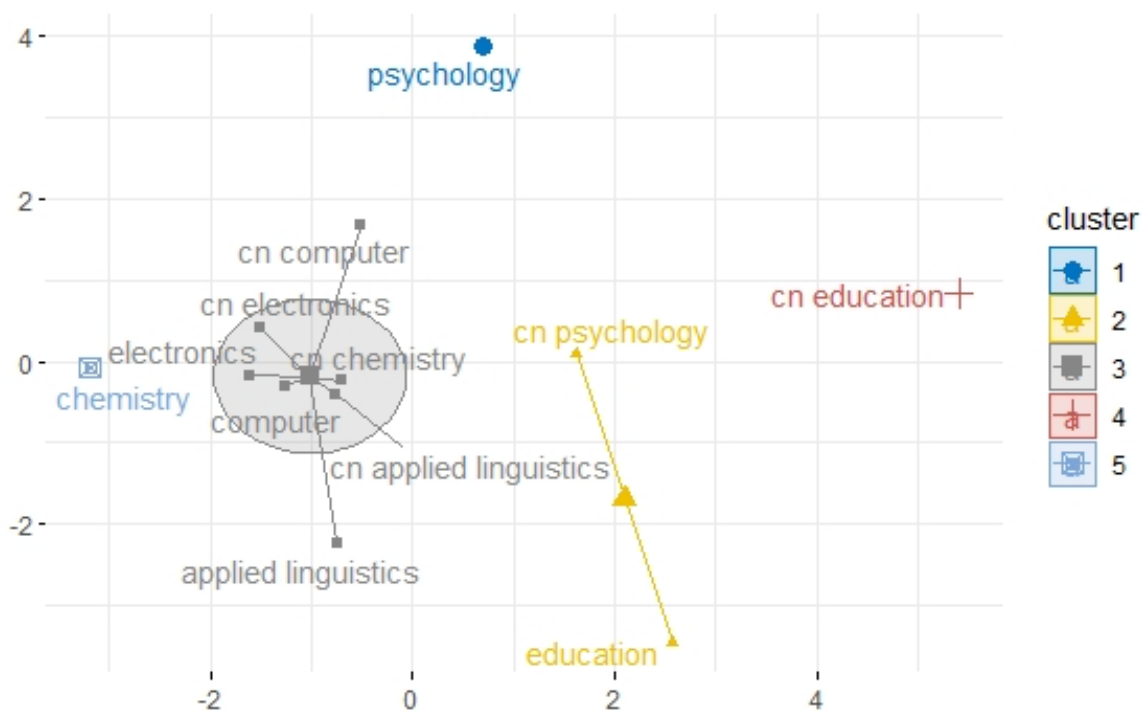


FIGURE 1. K-means clustering of EMs

Based on Figure 1, it is evident that the employment of EMs is more similar in hard sciences than in soft sciences for authors from the two linguistic backgrounds. As illustrated in Figure 1, the EMs in hard science disciplines such as computer, cn computer, electronics, cn electronics, and cn chemistry lie in the same cluster. For soft sciences, applied linguistics, and cn applied linguistics are also in the same cluster. This indicates that, generally speaking, the use of EMs in various subjects by English native and Chinese authors shows obvious similarities. There is, however, a difference in the use of EMs in the field of chemistry. To be more specific, as the figure shows, for English native and Chinese authors, the employment of EMs in the disciplines of chemistry and cn chemistry is diverse.

Interestingly, while there are some similarities in the use of EMs in cn psychology and education, as these two disciplines form parts of the same cluster, EMs are employed differently in the fields of psychology and cn education. As the figure shows, it can be concluded that in the fields of psychology and education, Chinese authors utilize EMs differently from English native authors.

EXEMPLIFYING MARKERS BASED ON LINGUISTIC BACKGROUNDS AND DISCIPLINES

The use of the above markers in the two linguistic backgrounds and the markers in the hard and soft sciences are shown in box plots for further investigation, as indicated in Figures 2 and 3.

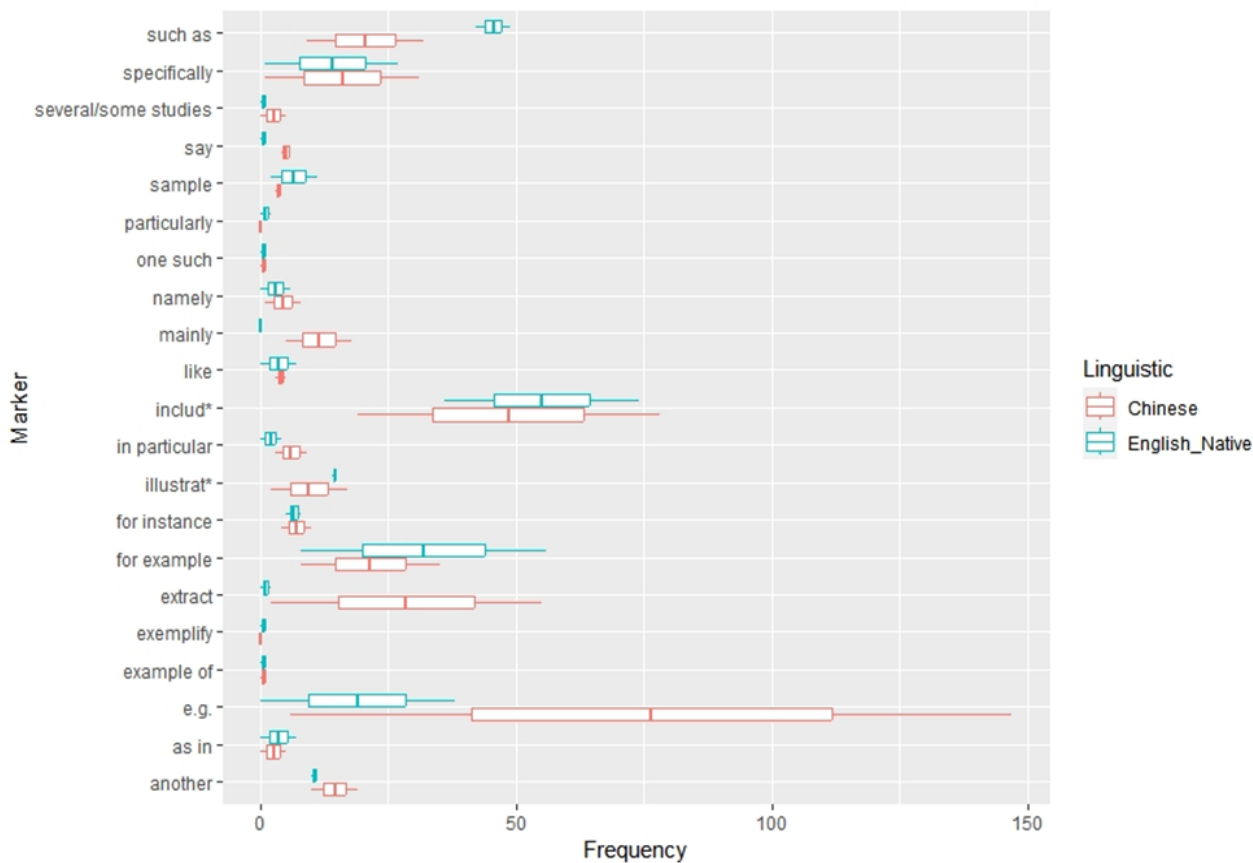


FIGURE 2. EMs in the two linguistic backgrounds

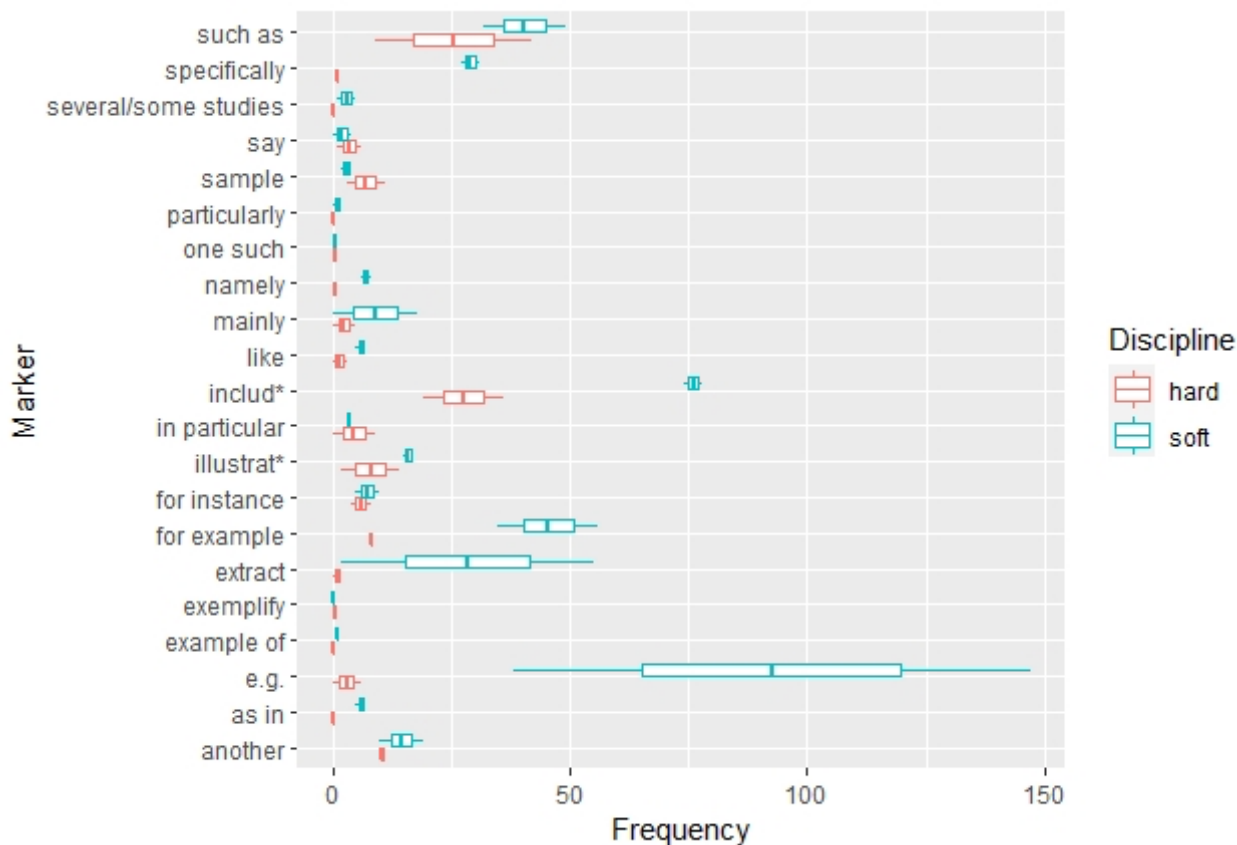


FIGURE 3. EMs in hard and soft sciences

Figure 2 shows that authors from different linguistic backgrounds employ EMs in different ways. Chinese authors tend to utilize certain markers, namely *e.g.*, *extract*, and *another* more frequently, while *includ**, *such as*, *for example*, and *illustrat** are more often chosen by English native authors. EMs are often employed more frequently in soft sciences than in hard sciences, as seen in Figure 3. In soft sciences, *e.g.*, *for example*, and *such as* are used more frequently than other EMs.

Welch Two Sample t-tests are carried out to identify comparisons based on the data. For articles in hard sciences and soft sciences written by English native authors, $t = -3.7947$, $df = 27.481$, and $p = .0007435$. Since $p < .05$, there is significant difference between English native authors using EMs in hard and soft sciences. The result of the Two Sample t-test for articles by Chinese authors in the hard sciences and soft sciences is, $t = -4.043$, $df = 23.105$, and $p = .0005019$; for English native and Chinese authors in the hard sciences is $t = -2.382$, $df = 37.698$, and $p = .02237$ and for English native and Chinese authors in the soft sciences, $t = -2.6702$, $df = 38.46$, and $p = .01105$.

All of the p-values in these data are less than 0.05, suggesting that there is a significant difference in the use of EMs in hard sciences and soft sciences RAs written by English native authors and Chinese authors. The difference is also notable between English native and Chinese authors in hard sciences as well as in soft sciences.

SEMANTIC TYPES OF EMS

EMs are categorized semantically based on the work of Rodriguez-Abrueiras (2015; 2019). According to the taxonomy, there are four types of EMs: neutral, hypothetical, comparative, and focalizing EMs. *For example, for instance* and *e.g.* are neutral EMs, *say* is a hypothetical EM, *like* and *such as* are comparative EMs, *especially, particularly, including, and included* are focalizing EMs. The framework is expanded with more examples from the list of EMs. Table 4 presents the semantic division of EMs by frequency.

TABLE 4. Semantic types of EMs
 (normalized to 10,000 words)

Types	Hard native	Soft native	Hard Chinese	Soft Chinese
Neutral EMs	26	91	22	121
Hypothetical EMs	0	0	3	4
Comparative EMs	24	53	23	40
Focalizing EMs	47	114	57	151

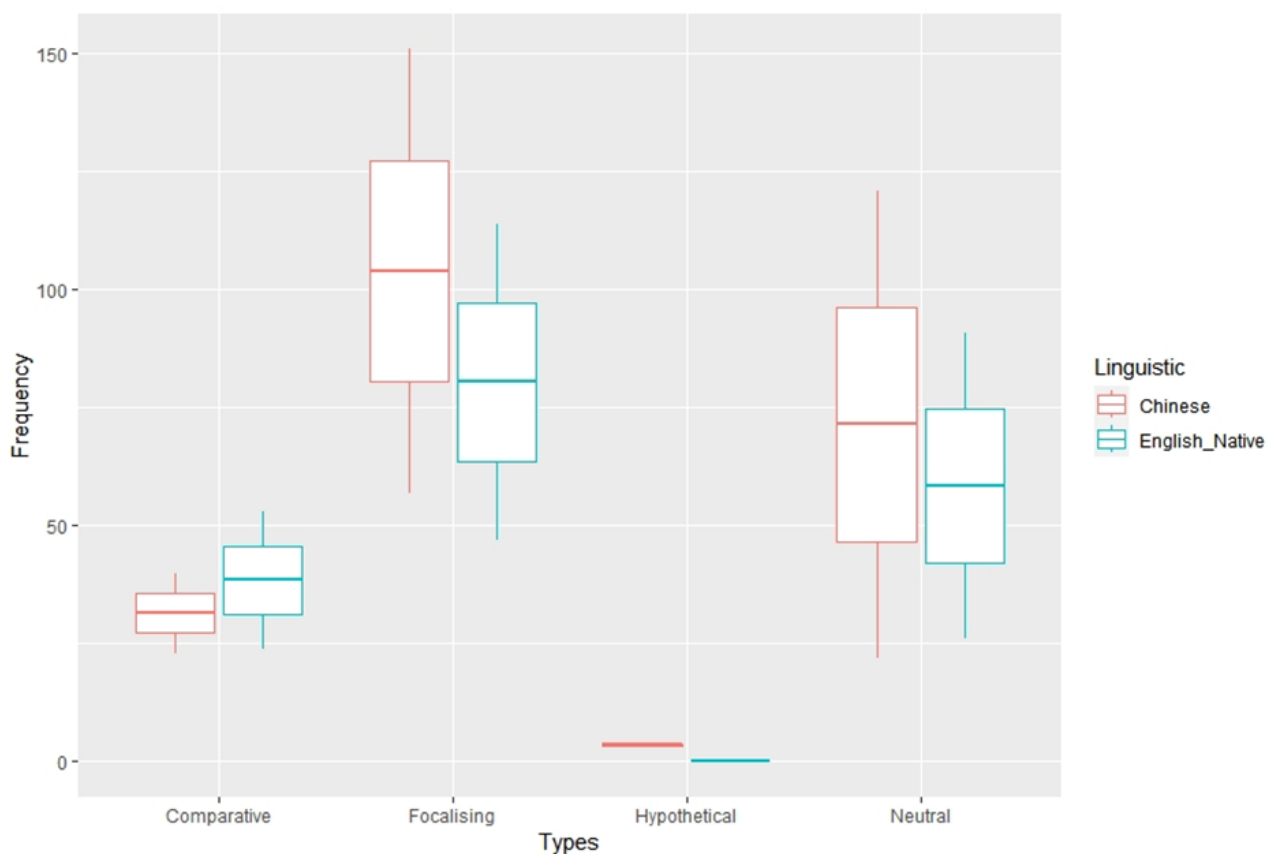


FIGURE 4. Semantic types of EMs in the two linguistic backgrounds

The semantic categories of EMs in articles produced by Chinese and English native authors are shown in Figure 4. Chinese authors use more focalizing EMs, neutral EMs, and hypothetical EMs than English native authors do. However, the difference is less pronounced. Compared to Chinese authors, native English speakers tend to use more comparative markers.

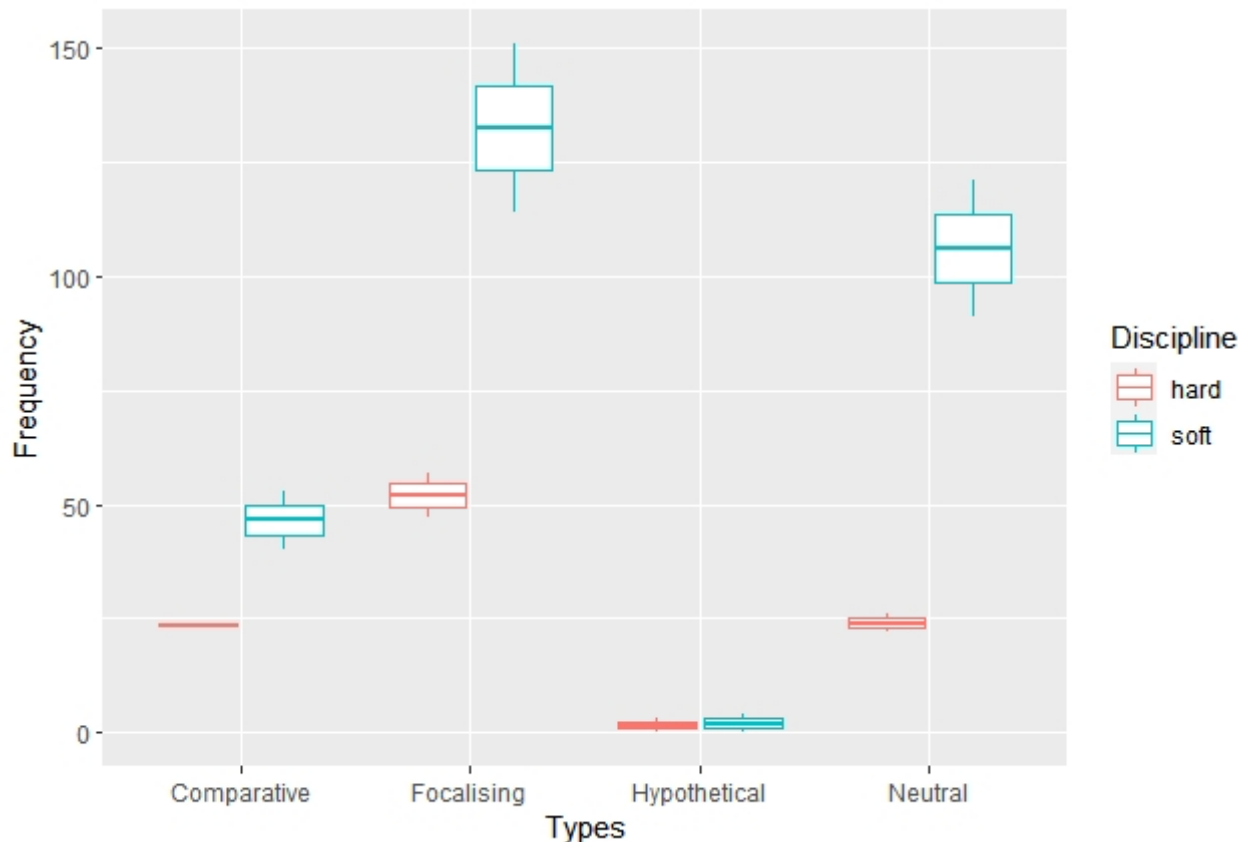


FIGURE 5. Semantic types of EMs in hard and soft sciences

Semantic categories of EMs in the hard and soft sciences are shown in Figure 5. The focalizing EMs are the most commonly employed type, followed by the neutral EMs and comparative EMs; the hypothetical EMs are utilized the least frequently in both hard and soft sciences, as indicated in the picture. As established, more markers are present in soft sciences than in hard sciences, particularly in the types of focalizing EMs, neutral EMs, and comparative EMs.

SYNTACTIC TYPES OF EMS

In addition to semantic types, there are three different syntactic types of exemplification based on the positions they appear (P1, P2, and P3). P1 is for the EM to be used before the Exemplifying Element (EE); P2 is for the EM to be used in the middle of the EE; and P3 is for the EM to be used after the EE. Please be noted that all the examples used in this paper are taken verbatim from the source; thus if there is an in-text citation in the example (as Example (1)), it remains as is, and it is also not listed in the reference of this paper.

Example (1) is an instance of exemplification in the P1 position.

(1) *Specifically*, we found that both “awareness of multiple layering in metaphors” (Low 1988: 134) and “producing (figurative) topic transitions/textual competence” (Littlemore and Low 2006a: 144-49) were associated with Productive Illocutionary MC. (Applied Linguistics)

In this instance, the EM, which is *Specifically*, comes before the EE, or the sentence that follows it.

The EM, *for example*, appears directly in the middle of the EE in Example (2) in P2 position.

(2) Well designed and enacted, partners and stakeholders all benefit from *the process by, for example, systematically codeveloping evidence-based solutions to address relevant problems of practice within a school* (Gutierrez & Penuel, 2014). (Education)

The EM that comes after the EE i.e. in the P3 position of exemplification is relatively uncommon, and it does not appear in the corpus.

EXEMPLIFYING UNITS

Exemplifying units serve as exemplification carriers and are the structures that immediately follow the EMs. These units can be nominal groups or clause types (see Figure 6). This type of structural differentiation is critical in understanding the relationship between units exemplified and units exemplifying.

Figure 6 shows that in all disciplines, more exemplifying groups than exemplifying clauses are used. The most exemplifying units occur in the discipline of education by English native authors, followed by the same discipline by Chinese authors. In the discipline of psychology by Chinese writers, there is a very high frequency of exemplifying units, and the next is applied linguistics by Chinese authors. In general, as expected, exemplifying units appear more frequently in soft sciences than in hard sciences RAs.

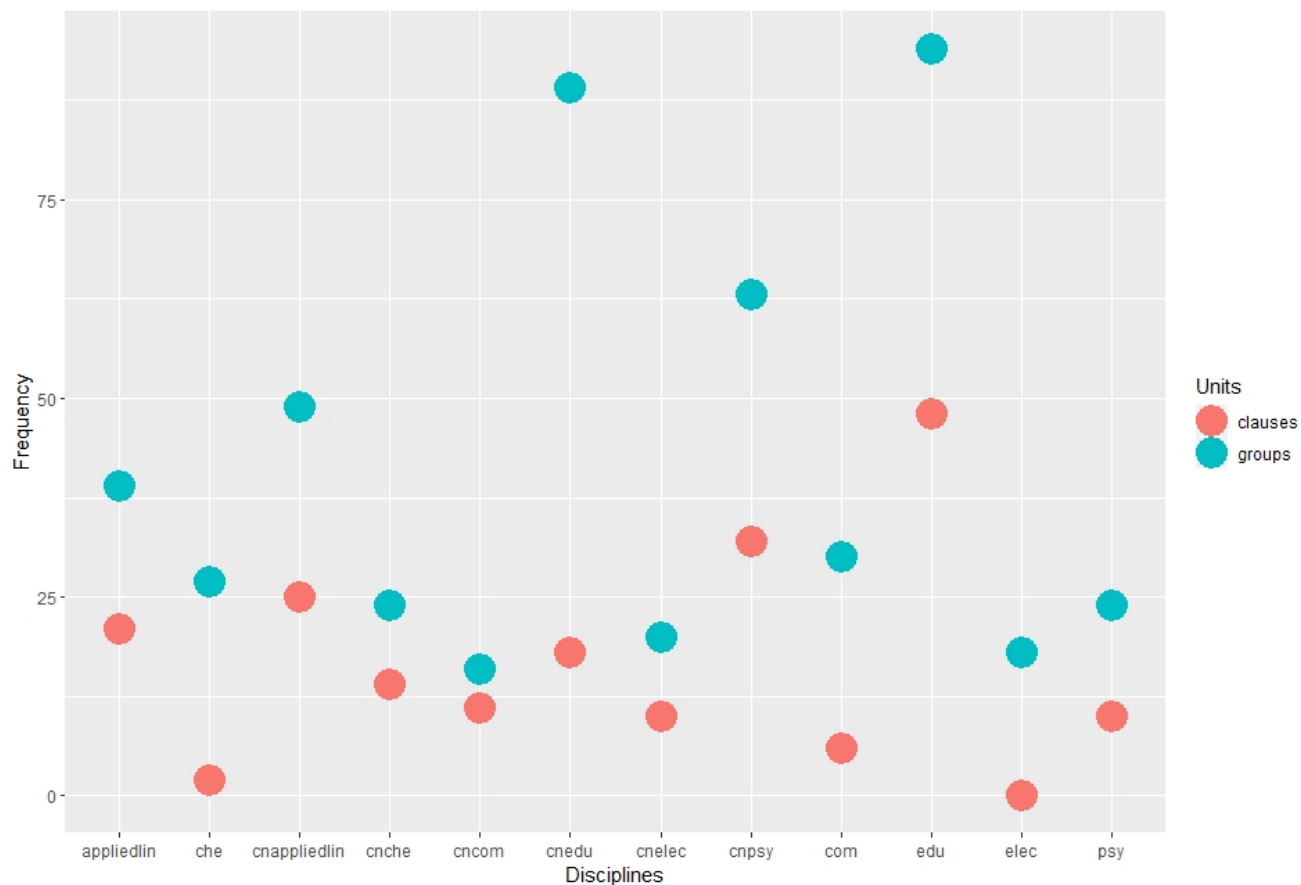


FIGURE 6. Exemplifying groups versus clauses in different disciplines

EXEMPLIFYING GROUPS

Exemplifying nominal groups can be simple, with minimal or no modification, or they can be complex, with varying degrees of modification (Triki, 2014). To help readers understand the concepts represented by the exemplifying unit, illustrative nominal groupings can represent entities that are either abstract or concrete (Triki, 2017).

(3) In order to solve the energy crisis and mitigate global warming issue, the renewable power generation systems, such as solar and wind power systems, have received a great interest. (cn Electronics)

Abstract notion *power generating systems* is explained in instance (3) with the exemplifying group that follow.

EXEMPLIFYING CLAUSES

Compared to exemplifying groups, exemplifying clauses have stronger elaborate power. The writers are given more space to elaborate on the ideas that the exemplified unit introduces (Triki, 2017). In contrast to the roles of nominal groups, examples, which present themselves as sentences, may offer readers more comprehensive and in-depth explanations. Exemplifying clauses benefit from their syntactical attributes, making meaning more transparent.

(4) More generally, identifying the effects of participant characteristics on L2 MC is also a challenge. For example, experimental studies have sometimes used fairly small groups of L2 learners with a wide range of L1s, which, given the cross-linguistic and cross-cultural differences in metaphor use. (Applied Linguistics)

Example (4) uses exemplifying clauses to further elaborate and clarify the concept *challenge*.

FUNCTIONS OF EXEMPLIFICATION

Exemplification serves a variety of purposes in RAs and analyses are carried out on functions of examples from the corpus. There are three different categories of exemplification functions, according to Hyland (2007). The first is to provide an example of a general category, for instance:

(5) L2 researchers have yet to tap into methods frequently used in L1 MC research, such as factor analysis, for exploring latent MC dimensions and their psychometric properties. (Applied Linguistics)

To demonstrate the point, an example of *L1 MC research, factor analysis*, is offered.

The second is to offer a parallel or similar case of the point.

(6) Drawing on Marxist, feminist, and/or critical theorists (Brown & Tandon, 1983; Brydon-Miller, 1997), pioneering PAR researchers boldly undertook social change, namely, human liberation, through various forms of inquiry (Freire, 1970; Horton, 1989; Maguire, 1987). (Education)

A case of social change, i.e., *human liberation* is provided following it to help readers better understand the concept of *social change*.

The third is to describe a precept or rule, for example:

(7) First, the FPSC nonlinear equation set may have no solution under some faulty situations by directly assigning Vani, Vni, and Vcni amplitude, such as 3-1-1 case (two cells in phases B and C are bypassed). (Electronics)

An example, *3-1-1 case (two cells in phases B and C are bypassed)*, is used for illustration of the precept, *situations by directly assigning Vani, Vni, and Vcni amplitude*.

Exemplification can be used for other purposes as well. Hedging operations allow it to function pragmatically and lessen the illocutionary force, as seen in Example (8).

(8) Visual strength *seems* to be determined by the low-level stimulus properties *such as spatial frequency components and contrasts*. (cn Psychology)

Exemplification is also an effective method for the transmission of knowledge and transfer of information. Example (9) accounts for the expression *macromolecular additives* with reference to a particular case, *hydroxypropyl cellulose (HPC)*.

(9) Finally, the incorporation of macromolecular additives, *such as hydroxypropyl cellulose (HPC)*, has been reported to red-shift the reflected colour of a CNC film while also suppressing edge inhomogeneities and acting as a plasticizer³². (Chemistry)

Barotto (2018b) asserts that examples can be utilized to clarify conceptual categories. In Example (10), studies that have *questioned the orientation's role as preattentive features* is stated with a specific case in history.

(10) However, several studies have questioned the orientation's role as preattentive features. *For example, Verghese and Nakayama (1993) showed that the performance on orientation stimulus drops substantially when there are more items, but there is almost no such drop for color stimulus*. (cn Psychology)

As demonstrated in Example (11), exemplification citations are used to enrich and explain the major point in the first sentence. Thus, exemplification can provide a link between sentences, which serves a complex rhetorical function.

(11) Other RPP researchers focused on how teachers made sense of equity and justice within their classrooms and curriculum. *For example, Santo et al. (2019) explored how K-12 educators conceptualized equity and how those conceptualizations shaped their planning and implementation of district-wide CS initiatives*. (Education)

In Example (11), *For example*, realizes its rhetorical purpose by acting as an inter-sentential link and joining two sentences.

With the introduction of the above examples, the general functions of exemplification in research articles can be understood, and a qualitative study of exemplification is realized.

DISCUSSION

This study is a corpus-based analysis of how Chinese and English native authors use exemplification when writing RAs in various disciplines. In the research of exemplification, this study focuses primarily on three aspects: EMs, exemplifying units and functions of exemplification.

Based on the findings, higher occurrences of exemplifications are found in the soft science RAs than in the hard science for both groups of authors, which resonates findings of previous studies: when the soft/hard distinction is considered, results show that examples are 14% more frequent in the soft sciences with 57% of all occurrences versus 43% in the hard ones (Hyland, 2007; Triki, 2014, 2017 and 2021).

Authors from the two linguistic backgrounds use EMs more similarly in hard sciences than in soft sciences. According to Schachter et al. (1991), the scientific method in the natural sciences is based on an accurate set of principles and regulated vocabulary that lacks room for synonyms. The hard disciplines operate in a knowledge context which has been established, and thus a specialized code is employed and the language is brief and exact (Schleef, 2008). Thus in hard sciences, EMs are more similarly utilized. In addition, K-means clustering analysis also reveals the difference in the use of EMs between Chinese authors and English-native authors in the fields of chemistry, psychology, and education. In scientific disciplines like chemistry, where adherence to precise scientific methods is crucial to ensure the experimental design, methodology, evaluation, interpretation, and dissemination of the outcomes are rigorous and impartial (Prager et al., 2019), English-native authors may employ EMs differently to support their arguments. However, the Chinese are often associated with high-context communication (Hall, 1976 as cited in Halford et al., 2018), characterized by their indirect and implicit communication. This cultural aspect may influence the use of EMs among the Chinese authors in the soft sciences domains of psychology and education, as well as in the field of chemistry.

Authors from the two linguistic backgrounds are also found to use EMs in various ways. Chinese authors typically use particular markers, e.g., *extract*, and *another*, but English native authors frequently choose *includ**, *such as*, *for example*, and *illustrat**. In addition, EMs like e.g., *for example*, and *such as*, are more frequently used in soft sciences than in hard sciences. Although Chinese authors use more focalizing, neutral, and hypothetical EMs than English native authors, the difference is not as pronounced. Native English speakers employ more comparative markers than Chinese authors. This finding may result from the direct and assertive communication commonly associated with the western culture, a clear difference from that of Chinese that advocates indirect argumentative method to build harmony and avoid confrontation, which is highly regarded in Confucian philosophy (Hinkel, 2002 as cited in Uysal, 2014). Thus, Chinese tend to use indirect ways to make comparisons and native English speakers use more comparative markers. In general, the focalizing EMs are the most frequently used type, followed by the neutral EMs, comparative EMs. Hypothetical EMs are used the least frequently in both hard and soft sciences domains.

More exemplifying groups than exemplifying clauses are employed across all disciplines. The most exemplifying units are found in the discipline of education written by native English speakers, followed by the same discipline written by native Chinese speakers. The frequency of exemplifying units in the discipline of psychology by Chinese authors is very high, followed by applied linguistics by Chinese authors. Exemplifying units are more common in soft sciences than in hard sciences in general. Moreover, different categories of functions are performed by exemplification in the corpus.

CONCLUSION

It is evident that authors from the two linguistic backgrounds employ EMs in hard sciences more similarly than in soft sciences. EMs are used in a variety of ways by authors from the two linguistic backgrounds. Chinese authors tend to employ specific markers, whereas English native authors tend to use others. Furthermore, EMs are more commonly used in soft sciences than in hard sciences. In hard sciences, researchers are interested in universal rules, quantities, and simplifications, and their work frequently leads to discoveries. While soft sciences are defined as holistic, concerned with details and quality, which result in interpretation and understanding.

Therefore, more EMs are applied in soft sciences. Some EMs, like *e.g.*, *for example*, *such as*, are more frequently used in soft sciences than the other markers.

Specifically, exemplifying markers in hard sciences and soft sciences RAs employed by each group of authors show significant differences. Similarly, significant differences in the use of exemplifying markers between English native and Chinese authors in hard sciences as well as soft sciences RAs are also observed. Authors with different linguistic backgrounds and disciplines prefer to utilize various categories of EMs. Furthermore, across the linguistic backgrounds and disciplines, more exemplifying groups than exemplifying clauses are used. The reason may be that exemplifying groups are simpler and easier to use.

This study can serve as a valuable guide for raising the level of diversity and enhancing idiomatic expression in the writing of Chinese authors. Additionally, it can also help Chinese students and scholars in enhancing their effectiveness in academic English writing. The theoretical significance of this study is as follows. Firstly, this study examines how Chinese authors and English native writers employ exemplification in English academic writing from a structural and functional standpoint, offering a fresh angle on the subject and developing new research concepts for academics. Secondly, this study investigates the similarities and distinctions in the use of exemplification in English academic writing between Chinese authors and English native writers. With this analysis, some relevant academic viewpoints can be improved, and the rhetorical theoretical system can be perfected.

Practically, by comparing the language habits of Chinese authors and English native writers in using exemplification in English academic writing, it is possible to identify the typical characteristics of Chinese authors in using interpretive language in English academic writing for different disciplines, as well as the language skills they have yet to master. The study can also aid in identifying the gap between Chinese and English native writers in English language writing. It is of tremendous practical importance to improve Chinese writers' academic writing and encourage their effective and correct expression of academic viewpoints in international academic interaction. Based on the findings, some implications can be offered for EAP education in China. Future studies should concentrate on a larger sample and more disciplines to have a more comprehensive understanding of exemplification in research articles.

REFERENCES

- Barotto, A. (2018a). The hedging function of exemplification: Evidence from Japanese. *Journal of Pragmatics*, 123, 24-37. <https://doi.org/10.1016/j.pragma.2017.09.007>
- Barotto, A. (2018b). The role of exemplification in the construction of categories: The case of Japanese. *Folia Linguistica*, 39(1):37-68. <https://doi.org/10.1515/flih-2018-0002>
- Barotto, A. (2021). *Exemplification and Categorization: The Case of Japanese* (Vol. 359). Berlin: De Gruyter Mouton.
- Bigsby, E., Bigman, C. A., & Martinez Gonzalez, A. (2019). Exemplification theory: A review and meta-analysis of exemplar messages. *Annals of the International Communication Association*, 43(4), 273-296. <https://doi.org/10.1080/23808985.2019.1681903>
- Crystal, D. (2003). *English as a Global Language*. Cambridge: Cambridge University Press.
- Dehghan, M. and Chalak, A. (2016). Code glosses in academic writing: The comparison of Iranian and native authors. *Research in English Language Pedagogy*. 3(2), 21-29. <https://www.researchgate.net/publication/303370650>

- Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in Contrast* (pp. 37-51). Lund: Lund University Press.
- Guziurová, T. (2022). Code glosses in L2 learner writing: Reformulation and exemplification in master's theses by Czech university students. In *Book of Abstracts of 6th Learner Corpus Research Conference*, pp. 52-53.
- Halford, W. K., Lee, S., Hiew, D. N., & van de Vijver, F. J. (2018). Indirect couple communication and relationship satisfaction in Chinese, Western, and Chinese-Western intercultural couples. *Couple and Family Psychology: Research and Practice*, 7(3-4), 183. <https://doi.org/10.1037/cfp0000109>
- Hall, E. T. (1976). *Beyond Culture*. New York, NY: Doubleday.
- Harvey, I. E. (2002). *Labyrinths of Exemplarity: At the Limits of Deconstruction*. Albany, NY: State University of New York Press.
- Hinkel, E. (2001). Giving personal examples and telling stories in academic essays. *Issues In Applied Linguistics (Los Angeles)*, 12(2), 149-170. <https://doi.org/10.5070/L4122005050>
- Hinkel, E. (2002). *Second Language Writers' Text: Linguistic and Rhetorical Features*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hu, C. (2001). *The Exemplification of Verbs in ECLDs and CELDs For Chinese Learners*. Unpublished Master's thesis, Guangdong University of Foreign Studies, Guangdong, China.
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse studies*, 7(2), 173-192. <https://doi.org/10.1177/1461445605050365>
- Hyland, K. (2007). Applying a gloss: Exemplifying and reformulating in academic discourse. *Applied linguistics*, 28(2), 266-285. <https://doi.org/10.1093/applin/amm011>
- Hyland, K. (2015). Genre, discipline and identity. *Journal of English for Academic Purposes*, 19, 32-43. <https://doi.org/10.1016/j.jeap.2015.02.005>
- Jasim, Y. B. (2023). Restating Research Findings in Research Articles Discussion Section: A Corpus Analysis of Linguistic Cues and Lexical Bundles. *GEMA Online® Journal of Language Studies*, 23(1), 225-243. <http://doi.org/10.17576/gema-2023-2301-12>
- Kafes, H. (2022). Communicative Functions of Code Glosses in Academic Discourse. *Selçuk Üniversitesi Edebiyat Fakültesi Dergisi*, (48), 53-72. <https://doi.org/10.21497/sefad.1218358>
- Krämer, B. & Peter, C. (2020). Exemplification Effects: A Meta-Analysis, *Human Communication Research*, 46(2-3), 192-221. <https://doi.org/10.1093/hcr/hqz024>
- Liu, X. (2017). Multimodal exemplification: The expansion of meaning in electronic dictionaries. *Lexikos*, 27(1), 287-309. <https://doi.org/10.5788/27-1-1404>
- Nesi, H., & Gardner, S. (2006). Variation in disciplinary culture: University tutors' views on assessed writing tasks. *British Studies in Applied Linguistics*, 21, 99.
- Paquot, M. (2008). Exemplification in learner writing: A cross-linguistic perspective. In F. Meunier & S. Granger (Eds.), *Phraseology in Foreign Language Learning and Teaching* (pp. 101-119). Amsterdam: John Benjamins.
- Petersen, C. (2020). What makes a good story?: Exemplification and explication of salient linguistic characteristics in a narrative preferred by the majority of a Danish population. *Pragmatics and Society*, 11(1), 24-44. <https://doi.org/10.1075/ps.17013.pet>
- Prager, E. M., Chambers, K. E., Plotkin, J. L., McArthur, D. L., Bandrowski, A. E., Bansal, N., ... & Graf, C. (2019). Improving transparency and scientific rigor in academic publishing. *Journal of Neuroscience Research*, 97(4), 377-390. <https://doi.org/10.1002/jnr.24340>

- Rahimpour, S. (2013). Employing a gloss in English and Persian academic writing: Exemplifying and reformulating in applied linguistics research articles. *Processing of the Global Summit on Education*, 749-763.
- Rodríguez-Abrunheiras, P. (2015). *Exemplifying markers in English: Synchronic and Diachronic considerations*. Ph.D thesis, University of Santiago de Compostela.
- Rodríguez-Abrunheiras, P. (2019). On exemplifying markers in present-day British and American English: formal and functional implications. *Brno Studies in English*. 45(2), 155-173. <https://doi.org/10.5817/BSE2019-2-8>
- Safari, I. (2018). A corpus-based contrastive study of code glosses used in English academic articles written by authors of politics and applied linguistics. *International Journal of Linguistics*, 10(2), 40-47. <https://doi.org/10.5296/ijl.v10i2.12242>
- Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology*, 60(3), 362-367. <https://doi.org/10.1037/0022-3514.60.3.362>
- Schleef, E. (2008). The “lecturer’s OK” revisited: changing discourse conventions and the influence of academic division. *American Speech*, 83(1), 62-84. <https://doi.org/10.1215/00031283-2008-003>
- Siepmann, D. (2005). *Discourse Markers across Languages: A Contrastive Study of Second-Level Discourse Markers in Native and Non-Native Text with Implications for General and Pedagogic Lexicography*. London & New York: Routledge.
- Su, H., & Fu, Y. (2023). Local grammar approaches to speech acts in Chinese: A case study of exemplification. *Journal of Pragmatics*, 212, 44-57. <https://doi.org/10.1016/j.pragma.2023.05.004>
- Su, H., & Lu, X. (2022). Assessing pragmatic performance in advanced L2 academic writing through the lens of local grammars: A case study of ‘exemplification’. *Assessing Writing*, 54, 100668. <https://doi.org/10.1016/j.asw.2022.100668>
- Su, H., & Zhang, L. (2020). Local grammars and discourse acts in academic writing: A case study of exemplification in Linguistics research articles. *Journal of English for Academic Purposes*, 43, 100805. <https://doi.org/10.1016/j.jeap.2019.100805>
- Su, H., Zhang, Y., & Chau, M. H. (2022). Exemplification in Chinese English-major MA students’ and expert writers’ academic writing: A local grammar based investigation. *Journal of English for Academic Purposes*, 58, 101120. <https://doi.org/10.1016/j.jeap.2022.101120>
- Su, H., Zhang, Y., & Lu, X. (2021). Applying local grammars to the diachronic investigation of discourse acts in academic writing: The case of exemplification in Linguistics research articles. *English for Specific Purposes*, 63, 120-133. <https://doi.org/10.1016/j.esp.2021.05.002>
- Taylor, Y. (2010). Functions of Japanese exemplifying particles in spoken and written discourse. Ph.D thesis, University of California, Los Angeles.
- Triki, N. (2014). Elaboration paradigms in PhD theses introductions. In M. Guirat, and M. Triki (Eds.), *Deviation (s)* (pp. 202-225). Tunisia: Laboratory on Approaches to Discourse (LAD).
- Triki, N. (2017). Elaboration across genre boundaries: Research articles and PhD dissertations as a case study. Ph.D thesis, University of Sfax, Sfax, Tunisia.
- Triki, N. (2021). Exemplification in research articles: Structural, semantic and metadiscursive properties across disciplines. *Journal of English for Academic Purposes*, 54, 101039. <https://doi.org/10.1016/j.jeap.2021.101039>

- Uysal, H. H. (2014). A cross-cultural study of indirectness and hedging in the conference proposals of English NS and NNS scholars. *Occupying Niches: Interculturality, Cross-culturality and Aculturality in Academic Research*, 179-195.
- Xu, H. (2008). Exemplification Policy in English Learners' Dictionaries. *International Journal of Lexicography*. 21(4), 395-417. <https://doi.org/10.1093/ijl/ecn015>
- Yuan, R. (2021). "Living in parallel worlds": investigating teacher educators' academic publishing experiences in two Chinese universities. *Compare: A Journal of Comparative and International Education*. 51(6), 787-805. <https://doi.org/10.1080/03057925.2019.1681260>
- Zillmann, D. (2002). Exemplification theory of media influence. *Media Effects: Advances in Theory and Research*. 2, 19-41.
- Zillmann, D., & Brosius, H. B. (2012). *Exemplification in Communication: The Influence of Case Reports on the Perception of Issues*. New York: Routledge.
- Zillmann, D., Gibson, R., Sundar, S. S., & Perkins Jr, J. W. (1996). Effects of exemplification in news reports on the perception of social issues. *Journalism & Mass Communication Quarterly*, 73(2), 427-444. <https://doi.org/10.1177/107769909607300213>

ABOUT THE AUTHORS

Yang Yuanyuan is a Ph.D student in Language Academy, Faculty of Social Sciences and Humanities, Universiti Teknologi Malaysia. She has participated and presented a paper in The International Conference on Social Sciences and Humanities 2022. Her areas of interest include Corpus Linguistics and Discourse Analysis.

Noor Mala Ibrahim is currently a Senior Lecturer in Language Academy, Faculty of Social Sciences and Humanities, Universiti Teknologi Malaysia. Her areas of research interest include academic writing, corpus linguistics, genre and applied linguistics.