A comparable-corpus-based study of informal features in academic writing by English and Chinese scholars across disciplines

Gao Xia
Beihang University, Beijing (China)
gaoxiasjtu@126.com

Abstract

This paper aims at quantitatively and qualitatively analyzing the use and distribution of informality features in a comparable corpus of research articles (RAs) written by L1 Chinese scholars (CSs) and L1 English scholars (ESs) across four disciplines. The normalized frequencies of eleven informal features were calculated and compared first between ESs' and CSs' RAs in the same discipline and then across the four disciplines. Four features, namely first person pronouns, pronominal anaphoric reference, sentence-initial conjunctions/conjunctive adverbs, and imperatives, were identified to be the contributing factors and analyzed qualitatively. The results demonstrate that: (i) there is significant difference in the use of informality features between ESs and CSs with CSs employing informality features less frequently than ESs; (ii) disciplinary variations are present with Physics RAs sounding more informal and Linguistics RAs more formal; (iii) the distribution of specific informality features presents a diversified picture: CSs' use of first person pronouns and pronominal anaphoric references is less frequent, and their use of imperatives and sentence-initial conjunctions/conjunctive adverbs is more frequent than ESs'. These findings shed light on teaching academic writing and provide writers with some guidance about stylistic choice.

Keywords: informality features, corpus, cross-disciplinary, academic writing, style.

Resumen

Estudio de los rasgos de informalidad en la escritura académica de investigadores ingleses y chinos de diferentes disciplinas en un corpus comparable

Este artículo analiza cuantitativa y cualitativamente el uso y la distribución de rasgos de informalidad en un corpus comparable compuesto por artículos de investigación de cuatro disciplinas escritos por dos grupos de investigadores conformados de acuerdo con su lengua materna: el chino o el inglés. Para ello, se han calculado las frecuencias normalizadas de once rasgos de informalidad y se han establecido comparaciones entre investigadores chinos y angloparlantes de la misma disciplina, por un lado, y entre las cuatro disciplinas, por otro. Se han identificado cuatro mecanismos especialmente relevantes, que han sido objeto de un análisis cualitativo: los pronombres de primera persona, la referencia anafórica pronominal, las conjunciones y adverbios conjuntivos en posición inicial, y los imperativos. Los resultados demuestran que (i) existe una diferencia significativa en el uso de rasgos de informalidad entre investigadores anglófonos y chinos, ya que los investigadores chinos los emplean con menor frecuencia; (ii) existen algunas variaciones entre disciplinas, dado que los artículos de Física resultan más informales que los de Lingüística; (iii) la distribución de algunos rasgos de informalidad específicos evidencia un panorama más heterogéneo: los investigadores chinos emplean con menor frecuencia los pronombres de primera persona y las referencias pronominales anafóricas, mientras que recurren con mayor frecuencia que los angloparlantes a los imperativos y a las conjunciones y adverbios conjuntivos en posición inicial. Estos resultados arrojan luz sobre la enseñanza de la escritura académica y ofrecen a los investigadores algunas orientaciones acerca de ciertas elecciones estilísticas.

Palabras clave: rasgos de informalidad, corpus, pluridisciplinar, escritura académica, estilo.

1. Introduction

A widespread general assumption in recent years is that writing in many domains has become less formal (e.g. Fairclough, 2001; Foster, 2005; Hyland & Jiang, 2017). Informality is said to have "invaded a large range of written and spoken domains once characterized by formality (journalism, business correspondence, administrative documents, etc.)" (Hyland & Jiang, 2017: 40). Whether this trend has spread to academic writing has become a growing interest of applied linguistics (e.g. Atkinson, 1999; Biber & Gray, 2016; Hyland & Jiang, 2017; Mair, 1998). Some diachronic corpus-based studies found a gradual shift from the detached, objective style to the more personal, involved trend in academic prose (Mair, 1998; Taavitsainen, 2002; Hyland & Jiang, 2017) while others found little influence of colloquialization on academic writing (Seone & Loureiro-Porto, 2005; Biber & Gray, 2016). Studies either focused on the use of particular lexical and grammatical

features regarded as informal such as first person pronouns (Flowerdew, 2001; Hyland, 2002; Harwood, 2005; Martínez, 2005; Mur Dueñas, 2007; Lafuente, 2010), imperatives (Swales et al., 1998), pronominal anaphoric references (Swales, 2005; Gray & Cortes, 2011; Wulff, Römer & Swales, 2012), contractions (Wang & Wang, 2017), and sentence-initial conjunctions (Bell, 2007) or lumped together all informal features and considered their total occurrence as evidence of more or less informality (Chang & Swales, 1999; Hyland & Jiang, 2017). Informality of academic writing was interpreted as containing spoken features (Barton, 1994; Lillis, 2013), deploying a narrative style (Biber, 1988), or using high percentages of verbs, prepositions, and pronouns etc. (Heylighen & Dewaele, 2002) in different studies.

While these investigations have shed light on professional writers' stylistic practices, few of them have focused on any particular writer groups and explored how informality is represented in their writing. That is, analyses available tell us little of how writers' linguistic backgrounds interact with disciplinary conventions to impact their stylistic choice in academic writing. In this paper we research how informality is represented in research articles (RAs) by L1 Chinese scholars (CSs) and L1 English scholars (ESs) across four disciplines. By investigating what types of informal features CSs and ESs use in their RAs and in what way their use differs, we hope to gain better understanding of the extent to which writers' linguistic backgrounds and disciplines influence their stylistic choice.

2. Prior research on the (in)formality of academic writing

Formality of writing has long been of interest in the study of language and discourse. It is difficult to define as readers claim a piece of writing as (in)formal based mainly on their instinct or general impression. Some researchers associated informality with everyday oral conversation in polar opposite to formal written language (Atkinson, 1999; Barton, 1994; Lillis, 2013) and defined spoken-like features such as involvedness, interaction, redundancy, false starts etc. as typical of informal style. Others, however, dismissed the modality distinction and defined formality according to the situation or context a discourse event occurs or the targeted audience or purpose (Heylighen & Dewaele, 2002). That is, formal discourse occurs

when one addresses an educated audience and needs to be precise, coherent, articulate, and convincing. Some studies on academic writing argue that (in)formality of a discourse is the occurrences of interwoven features prescribed by writing textbooks and journal style guides. Chang and Swales (1999), for example, in their survey of 40 style manuals, identified and investigated the use of 10 most commonly mentioned informal features in RAs and reported prominent use of some features (e.g., demonstrative pronoun *this*, sentence-initial conjunctions/adverbs) across disciplines and disciplinary preference for particular features (e.g., less contractions in Statistics). Hyland and Jiang (2017) furthered this line of study and investigated the use of 10 informality elements in 360 published RAs from 1965 to 2015 to find an increase of around 2% in the use of informality features with writing in hard science disciplines becoming more informal and that in social science slightly more formal.

Though residing at the slow end of "a cline of openness to innovation ranging from 'agile' to 'uptight' genres" (Hundt & Mair, 1999: 221), academic writing does change over time. Academics have paid close attention to this change and explore whether RA writing becomes more (in)formal. For example, Atkinson (1999), tracking changes in RAs published from 1675 to 1975, found that scientific writing in the early 20th century became less author-centered and more abstract and object-centered. Textual changes in the 21st century, nonetheless, were interpreted by Mair (1998), based on the Brown family corpora, as becoming more informal and colloquial. Similarly, Taavitsainen (2002) in her historical discourse analysis discovered a more personal or subjective trend in scientific discourse. Seone and Loureiro-Porto (2005), by comparing scientific English writing at three time periods, pointed out however that the oralization tendencies occur basically in the first half of the 20th century. Biber and Gray (2016) supported this finding and reported little influence of colloquialization on the discourse style of academic prose but strong influence on popular written registers based on a contrastive diachronic study on the use of colloquial features in written registers.

Other comparative studies investigated the use of informality features in learners' academic writing. Be they studies examining one category like writer/reader visibility (e.g., Petch-Tyson, 1998; McCrostie, 2008) or studies comparing a range of syntactic and lexical features (e.g., Shaw & Liu, 1998; Hinkel, 2003; Lee, Bychkovska & Maxwell, 2019), a similar conclusion was drawn, namely that learners' academic writing was infused with spoken,

informal language features. Among them, a study by Crawford (2005) found that German, Spanish and Bulgarian learners used more spoken features (personal pronouns, contractions, etc.) than native speakers, a study by Granger and Rayson (1998) reported that French learners overused many lexical and grammatical features typical of speech, and a study by Gilquin and Paquot (2007) provided further evidence for this claim reporting that learners of different backgrounds tend to use spoken features in their academic writing. More specifically, studies by Ma (2002), Pan (2012), Wang and Wang (2017), Wen (2009), and Xiao (2013) had Chinese learners, including undergraduate and graduate, English major and non-major students, as research targets and found their writing full of spoken features.

Both non-native and native novice writers are reported to lack register awareness and confuse or mix the use of informality features in their academic writing. An empirical corpus-based study into writers' stylistic choice in RA writing may reveal expert writers' practice and serve as reference for novice writers. Previous studies make no distinction between the stylistic choices made by different writer groups in RA writing. Further investigation is needed to explore how informality is represented in RAs by writers of different backgrounds and whether EFL writers' background impacts their stylistic choice. The present study intends to address this issue by investigating the distribution and use of informality features in English RAs by ESs and CSs across four disciplines. Following Hyland and Jiang (2017) and Chang and Swales (1999), we consider the total occurrence of informal features as the representation of RA informality and propose that the informality of a discourse is a concept of more or less rather than of yes or no. The selection of informality features and qualification for the selection will be discussed in section 3.2. Hopefully, a better understanding of writers' stylistic choice and its relationship with writers' backgrounds and disciplines will be gained to provide implications for EAP writing pedagogy.

3. Data and methods

3.1. The corpus

In this study we used the Beijing Collection of Academic Research Essays (Beijing CARE) as the data source. Compiled at the Department of Foreign Languages, Beihang University, the Beijing CARE is a new comparable corpus consisting of ESs' (26 million words) and CSs' RAs (11 million words) from

2010 to 2016 across 25 disciplines. RAs included in the ES and CS sub-corpus were downloaded from the same journals (the top seven regarding their rankings of impact factors) in each discipline from Elsevier databank. The first author of each article was identified as the main author, whose name and affiliation served as the classifying criteria. Based on and adapted from Wood (2001), "the first author with an English name and being at an institute in the inner circle of countries—Britain, America, Canada, Australia, and New Zealand was taken to be an ES; and the first author using Pinyin to romanize his/her Chinese name and based at universities in mainland China was taken to be a CS" (Gao, 2016: 16). Articles co-authored were included if the first author met the selection criteria so that proper quantity of articles can be downloaded to ensure suitable corpus size. This decision was made due to the limited number of CSs' single-authored articles (e.g., only 8 in the CS Computer sub-corpus).

The study reported here was based on a subsection of the Beijing CARE-RAs in four disciplines: Linguistics, Management, Physics, and Computer Science, selected as representing soft/pure, soft/applied, hard/pure, and hard/applied respectively based on Becher's (1989) taxonomy of disciplines. The detailed composition of this subsection is listed in Table 1. Based on the composition of the CS sub-corpora, we selected RAs in the same journal issue from the ES sub-corpora to make the corpus sizes of ES and CS sub-corpora in the same discipline comparable. That is, we included all the RAs in the CS sub-corpora and only part of RAs in the ES sub-corpora.

	Linguistics	Management	Physics	Computer	Total
CSs	368,159 (56)	300,674 (55)	328,467 (73)	316,074 (55)	1,313,374 (239)
ESs	367,744 (41)	304,909 (41)	327,070 (56)	316,022 (51)	1,315,745 (189)
Total	735,903 (97)	605583 (96)	655,537 (129)	632,096 (106)	2,629,119 (428)

Table 1. Sizes in tokens of the subsection (the number of RAs).

3.2. Informality features to be studied

Following Hyland and Jiang (2017), we calculated the sum of the normalized frequencies of informal features to represent the degree of RA informality. Eleven features of informality (see Table 2), ten of them selected and adapted from Hyland and Jiang (2017) and Chang and Swales (1999), were our research target as these features were commonly listed in textbooks (e.g., Swales & Feak, 2012: 22-25), frequently mentioned in style manuals and writing guidebooks (Chang & Swales, 1999: 147), and

researched in previous studies as representative of informal style (e.g., Hundt & Mair, 1999; Lee, Bychkovska & Maxwell, 2019). We, in accordance with the recognized associations between the pronominal use of demonstratives and informal style and the determiner use of demonstratives and formal style, limited the pronominal anaphoric reference to include *this*, *that*, *these* and *those* and excluded *it* mainly because it is impossible to determine when the use of pronoun *it* is informal as it cannot be used as a determiner and followed with a noun phrase. The imperative was added into the list for it is "an effective, although tricky, persuasive device" and has been "neglected by many scientific research writing and style manuals" (Swales et al., 1998: 3). "Neither the APA nor the MLA style manual seem to acknowledge the use of imperatives as a possible grammatical structure for the scholarly writer" (Swales et al., 1998: 4). Imperatives are nonetheless employed by many academics in their RAs to unleash themselves from an objective, impersonal style (Chang & Swales, 1999) and hence were included here.

	Features	Examples
1.	First person pronouns to refer to the author(s) (*I, we, me, us, my, our, mine,* and *ours*)	*In task 1, we examined the ideational meta-function of punctuations.*
2.	Unattended anaphoric pronouns (*this, that, these, those*) that can refer to antecedents of varying length	*This indicated that participants read sentences* for *comprehension*.)
3.	Split infinitives: an infinitive that has an adverb between *to* and the verb stem	*… to verbally describe the location of the objects…*
4.	Sentence-initial conjunctions or conjunctive adverbs	*And in the following section we will discuss …*
5.	Sentence-final prepositions	*This is the problem we try to deal with.*
6.	Listing expressions	*and so on, and so forth, etc.*, used when ending a list
7.	Second person pronouns/determiners to refer to the reader	*you, your*
8.	Contractions	*don't, can't, let's*, etc.
9.	Direct questions	*Is there any difference between the two groups*?
10.	Exclamations	*Someone should hire this man!*
11.	Imperatives	*See Table 4 for…, Suppose that…, Consider…*, etc.

Table 2. List of informal features.

3.2. Procedures

The corpus was POS-tagged using CLAWS 7.0 (Fligelstone, Pacey & Rayson, 1997), which has consistently achieved 96-97% accuracy. AntConc (Anthony,

2011) was used to carry out the quantitative analysis of informality features, which was supplemented with qualitative analysis as follows:

1) Some informality features, such as pronouns, listing expressions and exclamations, were retrieved easily by the software and then manually checked in context. As regards the first person pronouns, all the instances not referring to the author(s), or the discourse community were left out (such as *i* in *i.e.*, *us* abbreviating 'United States', *i* as an algebraic variable, etc.). Listing expressions when ending a list were picked out and saved for later analysis. Other features were extracted using a regular expression query based on their syntactic structure. All the features extracted were then checked manually line by line to ensure that they were a target feature. Instances within quotes or examples were deleted.

2) The frequencies of informality features were calculated and normalized at 10,000 words. Comparisons were made first between ESs' and CSs' RAs in the same discipline and then across the four disciplines.

3) Qualitative analyses concentrated then on the use of first person pronouns, pronominal anaphoric reference, sentence-initial conjunctions/conjunctive adverbs, and imperatives as they were identified to be the contributing factors to informality. Typical patterns were summarized and illustrated with examples.

4. Results and discussion

4.1. Use variations between writer groups

Table 3 shows that informal features appear frequently in both sub-corpora with more than 100 occurrences per 10,000 words, comparable to but less than the average 168-171 occurrences per 10,000 words reported in Hyland and Jiang (2017). The overall figures are largely influenced by the frequencies of four main features: first person pronouns, pronominal anaphoric reference, sentence-initial conjunctions/conjunctive adverbs and imperatives, which account for 94.86% to 98.16% of all informal elements across the sub-corpora. The results suggest that both CSs and ESs rely on similar informal features in RAs. This conforms to the finding reported in Hyland and Jiang (2017) that first person pronouns, sentence-initial conjunctions/adverbs, and unattended reference comprise the majority of informal features used in RAs. Despite the strictures stated in textbooks and style guides against the use of informal features (e.g., "avoid using *I*, *we*, and *you* in traditional academic

writing" (VanderMey et al., 2012: 79), unattended reference is "a common stylistic error" and always reflects "poor mechanical style" (Pfeiffer & Adkins, 2010:646), "don't use *but* at the beginning of a sentence in written English" (Longman Dictionary of Contemporary English Online)), professional writers seem to acknowledge the meta-discoursal functions these features realize in organizing text, constructing writer visibility and orienting readers and employ them frequently in RAs across disciplines.

Significant difference has been identified in the use of informality features in RAs by CSs and ESs as revealed by the results of the log-likelihood test (Log-likelihood=227.24, $p<0.001$). CSs are found to use less informality features in their RAs (CSs vs. ESs: 135.06 vs. 147.69 cases/10,000 words), which suggests CSs' proneness to closely follow the stylistic conventions compared with ESs' liability to "push gently at the boundaries of convention" (Casanave, 2010: 2). Although both groups rely on similar informal features, they differ in distinct ways. While CSs tend to employ more sentence-initial conjunctions/adverbs, imperatives and listing expressions, they utilize less first and second person pronouns, pronominal anaphoric reference, exclamations, direct questions, and split infinitives than ESs and appear to observe proscriptive rules more strictly. Given the rare occurrences (fewer than 5 items in any single L1/discipline cell of Table 3) and little impact of the seven informal features on the final scores (the totaling of all instances of all features), we will not elaborate these differences in turn. Qualitative analysis in section 4.3 will focus on the four contributing features.

	ES				CS			
	Ling	*Mana*	*Phys*	*Comp*	*Ling*	*Mana*	*Phys*	*Comp*
1st person pronouns	43.5	75.5	102.2	62.1	47.3	55.5	93.4	90.7
2nd person pronouns	0	0.72	0	1.93	0.46	0	0.03	0.13
Split infinitives	2.34	2.49	1.99	4.24	1.52	2.93	1.16	3.35
Listing expressions	0.027	1.12	0.489	0.759	1.44	0.566	0.792	2.63
Contractions (let's)	0	0.033	0.031	0.032	0.054	0	0.03	0.063
Sentence initial conjunctions and conjunctive adverbs	21.10	22.83	18.38	16.52	18.80	24.15	19.70	21.74
Pronominal anaphoric reference	42.937	42.373	44.822	40.662	28.140	31.296	24.660	30.594
Imperative	0.435	6.592	18.253	8.259	8.529	8.348	8.738	8.700
Sentence final preposition	0.408	0.394	0.489	0.570	0.299	0.333	0.457	0.190
Exclamation	0.027	0.033	0.397	0.127	0.027	0	0	0
Direct question	2.311	1.902	0.703	0.728	1.766	0.732	0.274	0.728
Total	**113.085**	**153.987**	**187.754**	**135.927**	**108.335**	**123.855**	**149.241**	**158.825**

Table 3. Normalized frequencies (*10,000 words) of informality features in CS & ES RAs.

4.2. Variations across disciplines

Cross-disciplinary variation in the use of informality features was found in both ES (Log-likelihood=686.23, $p<0.001$) and CS sub-corpora (Log-likelihood= 468.96, $p<0.001$). There is much higher use of informality features in hard science disciplines (Physics and Computer Science as illustrated in Table 3 in both CS and ES sub-corpora). For ESs, the average frequencies range from 187.75 cases per 10,000 words in Physics to only 113.09 cases in Linguistics. Informality features in CSs' RAs are particularly dense in Physics and Computer Science, but much less in Linguistics.

Previous studies nonetheless reported higher use of informality features in soft science disciplines, for example, in Philosophy in Chang and Swales (1999) and in Sociology in Hyland & Jiang (2017). Disciplines and informality features selected in these studies are different, which may account for the divergence in findings to some extent. Imperatives, the contributing factor identified in Chang and Swales (1999) and the present study, are not included in Hyland and Jiang (2017). Pronoun *it*, comprising a major share of use in Chang and Swales (1999) and Hyland and Jiang (2017), is excluded in the present study. Though all classified as representing hard science disciplines, Physics, Statistics, and Biology RAs seem to differ greatly in the use of informal elements with Physics RAs featuring high frequency of first person pronouns and imperatives, Statistics high frequency of imperatives, and Biology low frequency of pronouns. By employing *we* + *verb*, *our* + methodology- or result-related noun patterns and imperative sequences such as *Consider* ~, *See* ~, *Let* ~ *be* ~ etc. frequently, physicists make themselves visible in RAs and create active reader engagement. In (1), physicists use the *we* formula to present research findings and results and indicate their ownership of the findings, declare their responsibility for and commitment to the opinions.

> (1) From the plots in Fig. 15 *we see* that indeed there is now a significant fraction of very small weights, because many of the replicas fit the new data rather badly. (phys041101ra)

4.3. Variations in the use of specific features

First person pronouns, pronominal anaphoric reference, sentence-initial conjunctions/conjunctive adverbs and imperatives comprise the majority of informality features (see Table 3) and thus will be discussed in turn in this section.

First person pronouns, accounting for the largest share of informality features, have very different occurrences in ES and CS sub-corpora. ESs in Physics and Management employ first person pronouns significantly more frequently, while English computer scholars use them less frequently than their Chinese peers. The difference between ES and CS linguists is too slight to be significant, with more occurrences in CSs' RAs (47.3/43.5 cases/10,000 words respectively). This seemingly higher use might be explained by hypercorrection (Canagarajah, 2002) as Chinese linguists, in the course of their career experience, may observe the variation and thus hypercorrect to conform to the ES pattern. The higher use in RAs by Chinese Computer scholars is, nevertheless, difficult to explain and calls for further research. The null occurrence of self-reference using *I* in all eight single-authored RAs by Chinese computer scholars suggests that they are prone to present their work as a group using '*we* + verb' and '*our* + noun' patterns as they may consider the findings more accessible or acceptable to readers when using inclusive *we* and *our* (Gao, 2018).

Previous studies have reported similar less frequent use of first person pronouns in CSs' RAs (Hyland, 2002; Liu, 2011). Flowerdew (2001) and Martínez (2005) likewise report that Spanish scholars' English RAs feature lower authorial presence than ESs'. These findings suggest that EFL writers in general use first person pronouns less frequently in English RAs. That is, EFL writers seem to prudently hide behind their propositions to sound more objective and impartial whereas ESs are more willing to construct visible authorial presence in RAs.

Physics RAs in both ES and CS sub-corpora are prominent with the most frequent use (102.2/93.4 cases per 10,000 words). The frequent use of involved, interpersonal feature in Physics RAs contradicts the stereotype: the impersonal and objective style of hard science RAs stated in writing manuals. Similar usage patterns have been reported in Hyland (2001) on Physics and Biology and in Lafuente (2010) on Urology. This suggests that academics in some hard science disciplines favor the use of first person pronouns to modulate their relationship with their readers and the discourse community. "Self-mention might vary with different assumptions about the effects of authorial presence and rhetorical intrusion in different knowledge-making communities" (Hyland, 2001: 213). Disciplinary conventions exert a strong influence on RA writers' construction of authorial presence and identity as evidenced by the obviously more frequent use of first person pronouns in Physics RAs. The examples (2-5) extracted from the corpora illustrate how, by

strongly linking themselves to their findings, writers can solicit recognition for both:

(2) To maintain brevity, *we do not tabulate the results*; however, we find that the mean IPOUNPRC of our sample is 120%. (mana011201raCH)

(3) *We find* from Fig. 9(a) that at 08 BST on 20 March, the dust weather has occurred in the south part of Mongolia, inner Mongolia, Xinjiang, Gansu provinces and Beijing. (comp020401raCH)

(4) In fact, using traditional measures of elaboration –considering the use of dependent clauses– *we would conclude* that the opposite was the case: that conversation is more elaborated than academic writing. (ling051009ra)

(5) From Table 1, *we can see* that the Pearson 2x2 for time derivative of barycentric period, ……, drive the null hypothesis H0: the time derivatives of period and frequency of pulsars obey Benford's law, to be accepted. (phys051001raCH)

Pronominal anaphoric reference refers to the pronominal use of *this, that, these* and *those*, which, as a cohesive device, occurs frequently in academic prose (Swales, 2005). EAP textbooks and journal manuals, however, warn writers against the pronominal use and recommend instead the determiner use: the 'demonstrative + noun phrase' pattern (e.g., American Psychological Association, 2001; Lunsford, 2003; Swales & Feak, 2004). CSs in all four disciplines used fewer pronominal anaphoric references than ESs (Log-likelihood=31.15, $p<0.01$). Their far less frequent use might be explained by a teaching effect as EAP teachers "stressed the advantages of following *this* with a suitable summary/interpretive noun" (Swales, 2005: 8) to avoid the so-called unclear referent. Similar instruction can also be found in the APA style manual: writers are advised to 'eliminate ambiguity' by attending *this* with noun phrases (American Psychological Association, 2001).

Unattended *that* is the only sub-category used more frequently by CSs and therefore worth mentioning here. EFL such as Chinese, Polish and Swedish learners are also reported to use *that* as demonstrative pronoun more frequently than their native peers (Petch-Tyson, 2000; Leńko-Szymańska, 2004; Wang & Sun, 2006). CSs and learners "have clearly developed an awareness that distal demonstratives, particularly *that*, are less marked than proximal ones" (Leńko-Szymańska, 2004: 7) and they consciously or unconsciously opt for the unmarked form to shift reference focus or emphasize or evaluate an argument. When providing further explanation for

or comment on the previous statement, Chinese Physics and Computer Science scholars sometimes employ the pattern '*That means* …' as illustrated in (6-7). CSs seem to marginalize or other-attribute the proposition by using '*That means* …'. ESs more frequently use '*This means* …' to signal the referent of *this* "as the focus of attention for the subsequent text" (McCarthy, 1994: 274). The run-on sentence structure in (7) is original in the RA, which implies that the writer continues the information flow without recognizing the shift of focus by using '*that means* …'.

(6) Usually, non-linear statistical models outperform …. *That means* … (Comp031101raCH)

(7) The prerequisite for using … is that …, *that means* …. (Phys051203raCH)

Adverbs often occur in the mid-position of sentences in academic writing, while in informal English they are instead placed at the beginning or end of sentences (Swales & Feak, 2004, 2012). CSs are found to use more sentence-initial conjunctions/conjunctive adverbs in their RAs (84.39 cases/10,000 words) so as to build "an explicit linkage between two stretches of discourse" (Gao, 2016: 22). Tendency to front conjunctions/conjunctive adverbs to sentence-initial position has also been discovered in Japanese (Narita, Sato, & Suguira, 2004), Swedish (Altenberg & Tapper, 1998), French (Granger & Tyson, 1996), and Hong Kong (Field & Yip, 1992) EFL learners' academic writing. Both CSs and EFL learners attempt to ensure cohesive ties between sentences with explicit sentence-initial conjunctions/adverbs.

As the most frequent causal linking adverbial, *therefore* is used by CSs 89.5% of the time clause-initially (823 out of 920 occurrences). Some CSs "erroneously identify something as a cause simply because it occurs before the claim they are making or presenting" (Gao, 2016: 22). As illustrated in (8), we cannot derive any cause-and-effect relation between the two sentences. *Therefore* is superfluous as the logical relationship between the two sentences is expressed clearly by *i.e.* Another linking adverbial with significant use divergence between CSs and ESs is *meanwhile*. Its occurrence in ESs' RAs is rare (5 cases) compared to 85 instances in CSs' RAs. CSs almost invariably place *meanwhile* at sentence-initial position (82 out of 85 occurrences) to express, in addition to simultaneity, addition and contrast (39 out of 82 occurrences). *Meanwhile* in (9) may be substituted by *however*, as it points out the contrast between '5dB' and 'much smaller'.

(8) Now, let k = 0,1,2 ... [kmax]+1, and set the step size as $\Delta\varepsilon = 0.01$. *Therefore*, i.e., k = 0, 1, 2 ..., 92. (mana071202raCH)

(9) In the lowest two modes, the differences of TNLR between the proposed method and Ref2 is about 5 dB. *Meanwhile*, the absolute value of DSN for the proposed method is much smaller, which means ... (ling091302raCH)

Imperatives, though viewed as imposing face-threatening acts, appear in academic prose with varying frequency (Swales et al., 1998). The use of imperatives in RAs by ESs and CSs varies greatly (Log-likelihood=445.15, $p<0.001$). CSs in Linguistics, Management and Computer Science use imperatives more frequently and Chinese physicists use fewer than their English peers. When cross-disciplinary variations are investigated, we can see Physics in both ES and CS sub-corpora tops the rank. This corroborates Swales et al.'s finding that fields, "which tend to produce texts that not only consist of solid paragraph blocks, but also contain mathematical, experimental or illustrative elements […] may require rather more specific forms of reader-text management" (Swales et al., 1998: 102).

When the lexico-syntactic patterns of imperatives are summarized, we can see CSs and ESs employ similar patterns frequently: '*see* ~', '*note* ~', '*suppose* ~', '*consider* ~', '*let us* ~', '*let A Verb B*', '*assume* ~', '*notice* ~' and '*recall* ~'. Other imperative patterns and lexical choices occurring more than once are also summarized in Table 4.

As shown in Table 4, the '*see* ~' imperatives occur most frequently in our corpus. Two typical patterns are found, namely, '*see* ~' and '*for* ~ *see* ~'. The most frequent nouns occurring in these patterns are *figure, table, section, appendix* or *references*. Both ESs and CSs employ the *see* imperatives for two purposes: meta-discoursal to point the readers' attention to some sections in the paper or citational to cite relevant literature. Writers frequently employ the *see* imperatives to engage readers with the information flow by directing them to other parts of the text or relevant literature for further information. The shares of citational use are larger than those of inter-textual reference.

The second most frequent imperative pattern is '*note (that)* + clause', used as an attention-getting device. Both ESs and CSs tend to use this pattern "to draw the reader's attention to a crucial or unexpected argument […] by setting an emphatic tone" (Swales et al., 1998: 106). The occurrences of the *note* imperative are particularly dense in Physics. While Chinese physicists keep to the '*note that*…' form, some native physicists insert adverbs into the

	ES ling	ES mana	ES phys	ES comp	CS ling	CS mana	CS phys	CS comp
see ~	7	129	206	132	257	90	127	74
note ~	3	20	142	38	23	34	98	51
consider	0	17	72	15	3	19	2	7
suppose	1	5	6	17	7	22	6	35
let us	4	2	76	7	4	0	21	3
let A verb B	0	8	34	38	6	54	6	64
assume	0	5	3	2	1	7	2	22
notice	1	0	6	2	0	3	17	4
recall	0	11	18	0	0	4	2	2
take	0	0	8	0	2	1	1	4
define	0	0	0	1	0	3	0	1
put	0	2	0	0	3	0	0	0
observe	0	2	13	0	0	0	0	0
remember	0	0	3	2	1	0	0	0
presume	0	0	3	0	0	0	0	0
say	0	0	1	0	0	0	0	0
imagine	0	0	1	3	0	0	0	0
select	0	0	0	3	0	0	0	0
retrieve	0	0	0	1	0	0	0	0
use	0	0	0	1	0	0	0	0
look up	0	0	0	1	0	0	0	0
determine	0	0	0	2	0	0	0	0
develop	0	0	0	1	0	0	0	0
give	0	0	0	0	1	0	0	0
calculate	0	0	0	0	4	13	0	1
look at	0	0	0	0	1	0	0	0
créate	0	0	0	0	1	0	0	0
generate	0	0	0	0	1	0	0	0
compare	0	0	0	0	1	0	5	2
avoid	0	0	0	0	0	0	0	2
choose	0	0	0	0	0	0	0	1

Table 4. The occurrences of imperatives in the eight sub-corpora.

pattern as in '*Note, however, that*…', '*Note also that* …', and '*First, note that*…'. In addition to drawing readers' attention to an important point, writers also employ the '*note that*…' imperative to give further explanation. For example, '*Note that in the expression above we have denoted* …', which gives further explanation for the formula, and '*Note that C1(x) should be an even function of x* …', which demonstrates the difference between C2(x) and C1(x) further. Though infrequently (once only), both ES and CS computer scholars use the *note* imperative to compare the practice in the present study with that in relevant literature as in '*Note also that Hinckley et al. reported* …' and '*Note that although Horn et al. classified* …'.

The pattern '*suppose that* ~' is another one used in all four disciplines by both ESs and CSs albeit infrequently. CSs' frequency of use is higher than that in ES sub-corpora. CSs more often use the *suppose* imperatives in further explaining a calculus or formula as in '*Suppose L is the length of* …', '*Suppose R is a set on X * Y*…', and '*Suppose then that (A, u, v)S(A, u, v)S*'. They mention a possible situation for the formula using the *suppose* imperative, and then go on to discuss the potential effects this situation might have. In addition to giving further explanation of a calculus, Chinese computer scholars employ this imperative to begin the statement of a theorem or an example (e.g., '*Theorem 1. Suppose that (Gm, SIGm, VERm) is* …', '*Corolary 1. Suppose that* …', and '*Example 1. Suppose Alice and Bob are* …').

The *consider* imperative is also widely used by both ESs and CSs except L1 English linguists. Scholars commonly use the patterns such as '*consider* ~', '*consider for example* ~', and '*as an example, consider* ~' to further illustrate a point or discussion with exemplification. For example, L1 English computer scholars use '*As an example of this, consider Youtube*' to further illustrate the web 2.0 applications and Chinese Management scholars use '*Consider the numerical example which was discussed* …' to provide further explanation for the proposed approach in the previous sentence. While CSs' use of the *consider* imperative is lower and mainly relegated to the further illustration category, ESs' use is higher and has one more lexical pattern '*First/second/next/finally/now…, consider* ~', which serves as 'topic-initiator' to introduce a new topic. The co-selections 1) between a temporal or sequential adverb and 2) between the exemplification phrase and the *consider* imperative decide the different rhetorical functions being realized, that is, to start a new topic or to give further illustration. The information flow of the text either proceeds into the next stage as in the first pattern or is blocked as in the second one. CSs' neglect of the co-selection pattern may result in the far less 'topic-initiator' cases in their RAs. "The selection of different '*consider* ~' patterns not only reflects the rhetorical function of a certain expression, but may also signal to the reader something of the nature of the argumentation at this point in the author's text" (Swales et al. 1998: 105). Therefore, more exposure to real text is needed for CSs to sense the co-selection between patterns and their signaling functions.

The use of '*let us* ~' pattern is rare in other disciplines except in Physics. It is the most explicit invitation in all the imperatives and sounds less imposing than the direct '*Verb* ~' pattern. Both English and Chinese physicists use the '*let us* ~' imperative more often: in the beginning of a paragraph or a section

to introduce a new topic (e.g., '*First, let us briefly review* …', '*Let us begin with* …'), to make a suggestion by involving the readers into the action you are making (e.g., '*Let us take <φ0>≠0 <φ1>=0*', '*Let us define* …'), to elaborate the argument by giving examples (e.g., '*Let us take … as an example*', '*Let us consider … as the first example*'), or to direct the readers' attention towards the subject you want to consider next (e.g., '*Let us consider the set of poles* …'), or to "recapitulate previous examples to provide further evidence for an argument" (e.g., '*Let us recall that all vacua* …') (Swales et al. 1998: 12).

The illustration so far demonstrates that in academic writing scholars use imperatives, except for '*see* ~', mainly to "lead the readers through the argument" (Huddleston, 1971: 59). The use of the *see* imperatives is for citation and inter-textual reference, which serves a kind of text organization purpose. Some imperative patterns are used across disciplines and may be acceptable either for reader engagement or as tools for economy in academic writing. When using the explicit '*let us* ~' pattern, the authors tend to establish a closer link with the readers and make them feel that they are "on a par with the author" to be included in the discussion (Webber, 1994: 264), whereas the '*Verb* ~' imperatives sound more imposing like a command or request.

The distribution of the four contributing informality features discussed so far does not present a unanimous pattern, with CSs employing first person pronouns and pronominal anaphoric references less frequently and imperatives and sentence-initial conjunctions/conjunctive adverbs more frequently than ESs. That is, the distribution of specific informality features does not necessarily conform to the general pattern discovered based on the sum. We, in this study, calculate and add the frequencies of informality features and consider the sum as representing RA informality. It is worth considering whether the sum of parts equals the whole. Scholars may be invited to rate the degree of RA formality and the results of which might be compared with the total occurrences of the informality features to see if the sum of parts correlates with human rating. The finding here implies that the general tendency illustrated by the sum blurs the use difference of specific features.

5. Conclusions

This study, based on the comparable corpus of ES and CS RAs, has investigated the distribution of eleven informality features across four

disciplines and between two writer groups. The contrastive analysis revealed that CSs seem to stick to the proscription rules more strictly and employ less informality features in their RAs. Through instruction or observation in practice, CSs may be discouraged and avoid the use of informality features consciously to make their writing more formal and acceptable to journal editors.

The difference in use of informality features between ESs and CSs, however, must be heavily hedged as the identifying criteria for ESs and CSs are rather vague and the input of reviewers and editors has not been considered here. Though emails were sent to one-tenth of the first RA authors in the CS corpus to confirm their native speaker status, there might be some authors who meet the criteria but have received education abroad and therefore confound the data. We in this study assume that only the people listed as authors are responsible for the language used in RAs. The copy-editing process and journal reviewers' or editors' contribution are acknowledged but not included in our research focus. In future work interviews or a survey may be conducted to establish the influence of editors or other stakeholders in the writing process and on writers' stylistic choice.

Disciplinary variation in the use of informality features has been corroborated in the present study. RAs in hard science disciplines are found to contain more informality features, which deviates from the findings of previous research (Chang and Swales, 1999; Hyland & Jiang, 2017). Further discussion is needed for the verification of association between the delineation of soft and hard science disciplines and the use of informality features in RAs.

The selection of informality features in the present study is mainly based on previous literature (Swales et al., 1998; Chang & Swales, 1999; Hyland & Jiang, 2017). The validity of this method –lumping together the occurrences of all the features and calculating the sum as the representation of RA informality– needs further verification. As stated above, the sum blurs the use variation of particular features. Equal treatment of all informal features with no distinction in the degree of informality is also worth discussion as features such as an occasional exclamation seem to have greater impact on the reader than an additional use of *I*. How to quantify and represent the slippery concept of RA informality in a better way calls for further investigation.

A full picture of the potential use differences of informality features in academic writing between writer groups needs further contrastive analyses

covering writers of other linguistic backgrounds. Further complementary analyses, including qualitative studies, are needed to gather observations on how EFL scholars make strategic use of informal features in their RAs. The results reported here may have pedagogical implications as informality "may create additional complexities in the relationships the writer is seeking to build with readers and further increases to the compositional burden of novices" (Hyland & Jiang, 2017: 49). For example, imperatives, used to be considered as an informal feature, seem acceptable in RA writing as evidenced by their use frequencies in both ES and CS sub-corpora. Scholars across disciplines employ imperatives to draw readers' attention to a particular point or for engagement. "While style guides generally recommended authors to avoid them", these so-called informality features have been "legitimized in English academic writing" (Hyland & Jiang, 2017: 46). To what extent the use of informality features is acceptable in RAs calls for further investigation. Their occurrences in published RAs across disciplines may well imply gatekeepers' more liberal attitudes towards their use.

Acknowledgements

I am grateful to the two anonymous reviewers and the editor for their extremely valuable comments and suggestions. They have helped me enhance the quality of the paper significantly. All the remaining errors are my own responsibilities.

Article history:
Received 24 September 2019
Received in revised form 18 January 2020
Accepted 30 April 2020

References

Altenberg, B. & M. Tapper (1998). "The use of adverbial connectors in advanced Swedish learners' written English" in S. Granger (ed.), *Learner English on Computer*, 80-93. Harlow: Addison Wesley Longman.

American Psychological Association (2001). *Publication Manual of the American Psychological Association*. Washington, DC: American Psychological Association.

Anthony, L. (2011). *AntConc* (Version 3.4.2). Tokyo: Waseda University. Available from http://www.laurenceanthony.net/

Atkinson, D. (1999). *Scientific Discourse in Sociohistorical Context: The Philosophical Transactions of the Royal Society of London, 1675-1975*. Mahwah, NJ: Lawrence Erlbaum.

Barton, D. (1994). *Literacy: An Introduction to the Ecology of Written Language*. Oxford: Wiley-Blackwell.

Becher, T. (1989). *Academic Tribes and Territories: Intellectual Inquiry and the Cultures of Disciplines*. SRHE/OUP: Milton Keynes.

Bell, D. (2007). "Sentence-initial *And* and *But* in academic writing". *Pragmatics* 17(2): 183-201.

Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: CUP.

Biber, D. & B. Gray (2016). *Grammatical Complexity in Academic English: Linguistic Change in Writing*. Cambridge: CUP.

Canagarajah, S. (2002). *A Geopolitics of Academic Writing*. Pittsburgh: University of Pittsburgh Press.

Casanave, C. (2010). "Taking risks? A case study of three doctoral students writing qualitative dissertation at an American university in Japan". *Journal of Second Language Writing* 19: 1-16.

Chang, Y.Y. & J. Swales (1999). "Informal elements in English academic writing: Threats or opportunities for advanced non-native speakers?" in C. Candlin & K. Hyland (eds.), *Writing: Texts, Processes and Practices*, 145-167. London: Longman.

Crawford W. (2005). "Is L2 writing like native-English conversation?". Paper presented at *ICAME 26*, University of Michigan.

Fairclough, N. (2001). *Language and Power*, 2nd ed. Essex: Longman.

Field, Y. & L.M.O. Yip (1992). "A comparison of internal conjunctive cohesion in the English essay writing of Cantonese speakers and native speakers of English". *RELC Journal* 23: 15-28.

Fligelstone, S., M. Pacey & P. Rayson (1997). "How to generalize the task of annotation" in R. Garside, G. Leech & A. McEnery (eds.), *Corpus Annotation: Linguistic Information from Computer Text Corpora*, 122-136. London: Longman.

Flowerdew, L. (2001). "The exploitation of small learner corpora in EAP materials design" in M. Ghadessy, A. Henry & R. Roseberry (eds.), *Small Corpus Studies and ELT*, 363-379. Amsterdam: John Benjamins.

Foster, J. (2005). *Effective Writing Skills for Public Relations*, 3rd ed. London: Kogan Page.

Gao, X. (2016). "A cross-disciplinary corpus-based study on English and Chinese native speakers' use of linking adverbials in academic writing". *Journal of English for Academic Purposes* 24: 14-28.

Gao, X. (2018). "A cross-disciplinary corpus-based study on English and Chinese native speakers' use of first-person pronouns in academic English writing". *Text & Talk* 38(1): 93-113.

Gilquin, G. & M. Paquot (2007). "Spoken features in learner academic writing: Identification, explanation and solution" in *Proceedings of the 4th Corpus Linguistics Conference*, University of Birmingham.

Granger, S. & P. Rayson (1998). "Automatic lexical profiling of learner texts" in S. Granger (ed.), *Learner English on Computer*, 119-131. London/New York: Addison-Wesley/Longman.

Granger, S. & S. Tyson (1996). "Connector usage in the English essay writing of native and non-native ESL speakers of English". *World Englishes* 15: 17-27.

Gray, B. & V. Cortes (2011). "Perception vs. evidence: An analysis of *this* and *these* in academic prose". *English for Specific Purposes* 30: 31-43.

Harwood, N. (2005). "'We do not seem to have a theory… the theory I present here attempts to fill this gap': inclusive and exclusive pronouns in academic writing". *Applied Linguistics* 26(3): 343-375.

Heylighen, F. & J. Dewaele (2002). *Formality of Language: Definition, Measurement and Behavioral Determinants*. Internal Report, Center "Leo Apostel", Free University of Brussels.

Hinkel, E. (2003). "Simplicity without elegance: Features of sentences in L1 and L2 academic texts". *TESOL Quarterly* 37: 275-301.

Huddleston, R.D. (1971). *The Sentence in Written English: A Syntactic Study Based on an Analysis of Scientific Text.* Cambridge: CUP.

Hundt, M. & C. Mair (1999). ""Agile" and "Uptight" genres: The corpus-based approach to language change in progress". *International Journal of Corpus Linguistics* 4(2): 221-242.

Hyland, K. (2001). "Humble servants of the discipline? Self-mention in research articles". *English for Specific Purposes* 20(3): 207-226.

Hyland, K. (2002). "Authority and invisibility: authorial identity in academic writing". *Journal of Pragmatics* 34(8): 1091-1112.

Hyland, K. & F.K. Jiang (2017). "Is academic writing becoming more informal?" *English for Specific Purposes* 45(1): 40-51.

Lafuente, M. (2010). "'Extending this claim, we propose…' The writer's presence in research articles from different disciplines". *Ibérica, Journal of the European Association of Languages for Specific Purposes* 20: 35-56.

Lee, J.J., T. Bychkovska & J.D. Maxwell (2019). "Breaking the rules? A corpus-based comparison of informal features in L1 and L2 undergraduate student writing". *System* 80: 143-153.

Leńko-Szymańska, A. (2004). "Demonstrative as anaphora markers in advanced learners' English" in G. Aston, S. Bernardini & D. Stewart (eds.), *Corpora and Language Learners*, 89-107. Amsterdam/ Philadelphia: John Benjamins.

Lillis, T. (2013). *The Sociolinguistics of Writing*. Edinburgh: EUP.

Liu, S. (2011). "Self-mentions and authorial identity construction in Chinese and native scientists' abstract". *Modern Rhetoric* 166(4): 85-88.

Lunsford, A.A. (2003). *The St. Martins Handbook*, 5th ed. Boston: Bedford St. Martins.

Ma, G.H. (2002). "Contrastive analysis of linguistic features between EFL and ENL essays". *Foreign Language Teaching and Research* 34(5): 345-349.

Mair, C. (1998). "Corpora and the study of the major varieties of English: Issues and results" in H. Lindquist, S. Klintborg, M. Levin & M. Estling (eds.), *The Major Varieties of English: Papers from MAVEN 97, Växjö*, 139-157. Växjö: Edman and Westerlunds Tryckeri.

Martínez, I. (2005). "Native and non-native writers' use of first person pronouns in the different sections of biology research articles in English". *Journal of Second Language Writing* 14(3): 174-190.

McCarthy, M. (1994). "*It*, *this* and *that*" in M. Coulthard (ed.), *Advances in Written Text Analysis*, 266-275. London: Routledge.

McCrostie, J. (2008). "Writer visibility in EFL learner academic writing: A corpus-based study". *ICAME Journal* 32: 97-114.

Mur Dueñas, P. (2007). "'I/we focus on…': a cross-cultural analysis of self-mentions in business management research articles". *Journal of English for Academic Purposes* 6(2): 143-162.

Narita, M., C. Sato & M. Sugiura (2003). "The use of linking adverbials in the English essay writing of Japanese EFL learners" in M. Murata, S. Yamada & Y. Tono (eds.), *ASIALEX '03 Tokyo Proceedings*, 440-444. Tokyo: Asian Association for Lexicography.

Pan, F. (2012). "MF and MD analysis of written texts produced by Chinese non-English major undergraduates and graduates". *Journal of Foreign Language Teaching and Research* 44(2): 220-232.

Petch-Tyson, S. (1998). "Writer/reader visibility in EFL written discourse" in S. Granger (ed.), *Learner English on Computer*, 107-118. London: Longman.

Petch-Tyson, S. (2000). "Demonstrative expressions in argumentative discourse: A computer corpus-based comparison of non-native and native English" in S.P. Botley & A.M. McEnery (eds.), *Corpus-based and Computational Approaches to Discourse Anaphora*, 43-64. Amsterdam/ Philadelphia: John Benjamins.

Pfeiffer, W.S. & K.E. Adkins (2010). *Technical Communication: A Practical Approach*, 7th ed. Upper Saddle River, NJ: Longman.

Seone, E. & L. Loureiro-Porto (2005). "On the colloquialization of scientific British and American English". *ESP Across Cultures* 2: 106-118.

Shaw, P. & E. Liu (1998). "What develops in the development of second-language writing?" *Applied Linguistics* 19: 225-254.

Swales, J.M. (2005). "Attended and unattended "this" in academic writing: A long and unfinished story". *ESP Malaysia* 11: 1-15.

Swales, J.M., U. Ahmad, Y. Chang, D. Chavez, D. Dressen & R. Seymour (1998). "Consider this: The role of imperatives in scholarly writing". *Applied Linguistics* 19: 97-121.

Swales, J.M. & C.B. Feak (2004). *Academic Writing for Graduate Students: A Course for Nonnative Speakers of English*, 1st ed. Ann Arbor: University of Michigan Press.

Swales, J.M. & C.B. Feak (2012). *Academic Writing for Graduate Students: A Course for Nonnative Speakers of English*, 3rd ed. Ann Arbor: University of Michigan Press.

Taavitsainen, I. (2002). "Historical discourse analysis. Scientific language and changing thought-styles" in T. Fanego, B. Méndez-Naya & E. Seone (eds.), *Sounds, Words, Texts and Change*, 201-226. Amsterdam/ Philadelphia: John Benjamins.

VanderMey, R., V. Meyer, J. Van Rys & P. Sebranek (2012). *The College Writer*, 4th ed. Boston, MA: Wadsworth.

Wang, L.F. & X.K. Sun (2006). "A corpus-based study on demonstratives in argumentative discourse of Chinese EFL learners". *Modern Foreign Languages* 29(2): 172-179.

Wang, L. & N. Wang (2017). "Colloquialisation in EFL learners' theses in China". *Modern Foreign Languages* 40(2): 275-286.

Webber, P. (1994). "The function of questions in different medical journal genres". *English for Specific Purposes* 13: 257-268.

Wen, Q. (2009). "A study of changes of register features in learner English". *Journal of Foreign Languages* 4: 2-10.

Wood, A. (2001). "International scientific English: The language of research scientists around the world" in J. Flowerdew & M. Peacock (eds.), *Research Perspectives on English for Academic Purposes*, 71-83. Cambridge: CUP.

Wulff, S., U. Römer & J. Swales (2012). "Attended/unattended this in academic student writing: Quantitative and qualitative perspectives". *Corpus Linguistics and Linguistic Theory* 8(1): 129-157.

Xiao, F. (2013). "An investigation into how oral/written register features relates to grade levels and ESL writing quality". *Foreign Languages in China* 10(6): 65-71.

Gao Xia is an Associate Professor of English at Beihang University in China. She earned her PhD in Linguistics at Shanghai Jiao Tong University in 2006. Her research interests cover Corpus Linguistics, English for Academic Purposes and Second Language Acquisition.