



Contents lists available at ScienceDirect

Journal of English for Academic Purposes

journal homepage: www.elsevier.com/locate/jeap



However and other transitions in the Han CH-EN corpus

Chao Han^a, Sheena Gardner^{b,*}

^a Lanzhou University, China

^b Coventry University, UK



ARTICLE INFO

Article history:

Received 21 September 2020

Received in revised form 27 February 2021

Accepted 1 March 2021

Available online 5 March 2021

Keywords:

Metadiscourse

Academic writing

Transitions

BAWE

Chinese

Overuse

BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Transitions are numerous when compared with other metadiscourse resources in academic writing (Hyland, 2005). This suggests that they are important resources for students to control, and they are often a focus in teaching English for academic purposes. Transitions have also been widely studied both as part of metadiscourse (e.g. Ho & Li, 2018) and more generally (e.g. Biber et al., 1999), and although each study has contributed to our understanding of specific aspects of transitions in certain contexts, a comparison of the findings across studies raises contradictions and inconclusive claims.

Specifically, contrastive studies of Chinese and English writers in English have noted that certain transitions are overused or underused by Chinese writers. One aim of our study is to explore whether such claims stand up when they are examined in a closely matched corpus of Chinese and English writing.

Secondly, contradictory claims and explanations have been made about the discipline and genre¹ of texts where transitions are more frequent. For example, Biber et al. (1999) suggest transitions are most frequent in technical scientific writing, while Hyland (2005) suggests they are more frequent in the discursive writing of the soft disciplines. A second aim of our study is to address these contradictions with reference to the most frequently used transition in our data, *however*.

* Corresponding author.

E-mail address: ab1248@coventry.ac.uk (S. Gardner).

To avoid the pitfall of comparing Chinese texts from one context with English texts from a different context, we examine transition use in the matched Han CH-EN corpus. Our assumption is that an investigation of transitions in a closely matched corpus will prove invaluable in resolving contradictory or inconclusive findings for teachers and learners of English for academic purposes. In particular, this paper will not only identify all transitions used by Chinese and English students in the successful academic writing of the Han CH-EN corpus to initially explore the 'over/under use' questions, but also examine the frequent transition, *however*, in its contexts of use across genres, disciplines and syntactic structures to explore the competing claims around transitions in technical scientific vs discursive writing in the soft disciplines.

2. Transition use in English by writers with different language backgrounds

There is evidence to suggest that 'good' writers, whether L1 English or ESL/EFL, use a wider range of transitions and use them more densely (as reviewed in [Knoch et al., 2014](#)). This is not supported by all studies of university student writing.

In a comparison of essays from university students taking Arts degrees in England and in France, [Granger and Tyson \(1996:21\)](#) found overall no substantial difference in the frequency of use of connectors in English essay writing between native English and advanced EFL French speakers, while [Kennedy and Thorp \(2007\)](#) found that proficient writers at IELTS 8 used markers like *however* less frequently than weaker writers at levels 4 and 6. This suggests that at university level, lower proficiency writers may rely more heavily on transitions to guide their readers than higher proficiency writers.

Granger and Tyson also noted differences in the sentence position of connectors. For example, Chinese learners typically used *however* in sentence-initial position (87%), while their English counterparts distributed it more equally between sentence-initial (49.7%) and non-initial positions ([Granger & Tyson, 1996:25](#)).

More recently, the British Academic Written English (BAWE) corpus has been used in contrastive Chinese-English studies of successful university student writing. [Leedham \(2012\)](#) examined a BAWE subcorpus of 29 texts (48,367 words), submitted by four university students, who had completed high school in China. She compared their texts from years 1/2 and year 3 of undergraduate study. Three discipline-matched reference subcorpora of L1 English BAWE texts were also compiled. She suggested that Chinese students overuse some connectors (e.g., *on the other hand*), which is consistent with [Chen and Baker \(2010\)](#), though neither produced an exhaustive account of the connectors examined.

In terms of sentence position, [Leedham \(2015:46\)](#) using a mixed corpus with texts from BAWE and other sources ([Leedham, 2015:8](#)) found that Chinese undergraduate writers preferred (88%) sentence-initial pattern for *however*, as did English students (65%). She also identified that overused connectors in Chinese students' assignments, such as *on the other hand* and *in the long run*, tended to be used sentence-initially. This is in broad agreement with previous studies (e.g. [Granger & Tyson, 1996](#); [Lee & Chen, 2009](#)).

These studies suggest that successful Chinese university students do prefer a more limited set of transition markers in a more limited range of positions. While in some cases this may be a strategically safe choice, it may also result from the lists of transitions provided in textbooks and dictionaries which are insufficiently differentiated for use in different contexts ([Lee & Chen, 2009](#); [Milton, 1999](#)).

3. Transitions in written academic genres and disciplines

Compared with other metadiscourse resources, transitions are numerous and dense in written academic texts. This makes them not only important in revealing how writers guide their readers through arguments in the text, but also amenable to study across written academic genres and disciplines.

In studies of university textbooks, [Hyland \(2005:102\)](#) found that transitions accounted for around 40% of all metadiscourse items and almost 60% of interactive items. Similarly, [Biber \(2006:70\)](#) found that the most common linking adverbials², – *however* and *for example* – occurred around once per thousand words in textbooks, where they were more frequent than in course management and institutional writing texts.

Transitions are also frequent in student writing, though relatively less so than in textbooks. Studies of graduate student writing in Hong Kong showed a high proportion of transitions, with hedges also frequent. In PhD student writing, transitions and hedges were the most frequent features, at 95.6 per 10,000 words whereas in Masters student texts, they were second to hedges at 75.8, accounting for approximately 40% of the interactive metadiscourse ([Hyland, 2005:55](#)).

In comparing studies of transitions in textbooks and in research articles, Hyland found that they accounted for 41% of all metadiscourse in textbooks, which is more than double the 19% they accounted for in research articles (2005:102). One reason for this is that hedging is proportionally more frequent in research articles than in student textbooks. When frequencies are normalized, moreover, we see that transitions were more than twice as frequent (28.1 vs 12.3 per 1000 words) in textbooks than research articles. This suggests that more explicit guidance is given to students than to fellow researchers.

Not all transitions behave in the same way, however. [Yin's \(2014\)](#) investigation of patterns of form, meaning and position of linking adverbials across registers reveals that some, like *however*, are most frequent in written academic prose where they also retain a constant (adversative) semantic meaning, whereas others, like the discourse marker *so*, can carry up to nine different pragmatic meanings in spoken lectures (e.g. restatement, result, conclusion). These findings all support the argument that the use of metadiscourse features varies consistently with the different audiences, purposes and contexts of different academic texts.

In a study of textbooks across five disciplinary groups, 'Business and Engineering textbooks show the greatest reliance on linking adverbials.' (Biber, 2006:71). In studies of graduate student writing, Hyland (2005:57) suggests that there are more metadiscourse features in the discursive 'soft' sciences than in the 'hard' sciences. Similarly, in his textbook study, transitions occur around 95, 98 and 89 times per 10,000 words in the more discursive Applied Linguistics, Public Administration and Business Studies, compared to around 74, 77 and 87 in Computer Science, Electronic Engineering and Biology, respectively. Surprisingly perhaps, this difference does not carry through to research articles, where transitions occur to a similar extent: 111 times per 10,000 words in Applied Linguistics and 113 times in Biology (Hyland, 2005:144). In contrast, Biber suggests it is the hard disciplines that contain more transitions: 'In both disciplines [Business and Engineering], linking adverbials are most common in technical discourse that explains the derivations or application of mathematical formulas or procedures' (Biber, 2006:71), which is the sort of discourse we might also expect in Computer Science and Electronic Engineering. This means that we have one study saying that transitions are more frequent in the soft, discursive sciences, one study saying that there is no substantial disciplinary difference, and a third study suggesting that linking adverbials are more frequent in the hard sciences.² We will contribute to this debate by examining the transition use across genres and disciplines in university student writing.

Our review also shows that transitions are widely and densely used, their use is differentiated by genre and by discipline, while the behavior of individual transitions varies, and full details are not yet understood.

4. Methodology

In order to provide answers to questions about the frequency and distribution of transitions in the writing of Chinese and English students, we first compiled the Han CH-EN corpus (4.1), identified transitions in the corpus (4.2), and compared them in terms of frequency and statistical difference (4.3). There follows quantitative and more qualitative comparisons of the use of *however* which emerged as not only the most frequent transition for both groups of writers, but also a transition that unusually was used statistically more frequently by English students.

4.1. The Han CH-EN corpus

The Han CH-EN corpus, which is a subcorpus of the British Academic Written English (BAWE) corpus, was developed to enable the comparison of transition use by Chinese and English university student writers. The texts are all individually (i.e. not group) authored, and successful, in that they have been given top grades in regular coursework. The Chinese students selected all self-declared their first language (L1) as 'Chinese', 'Mandarin' or 'Cantonese', and were not educated in the UK prior to university. The English students selected all self-declared English as their first language and were fully educated in the UK prior to university.

The 78 pairs of texts in the Han CH-EN corpus are highly-matched in that for each text by a Chinese student, there is one from the same discipline, genre family and level of study by an English student. Where possible, pairs of texts are from the same module assignment. For example, Chinese student's text 0008a matches English student's text 0354b in that both are Case Study assignments from a first year undergraduate Engineering module 'Economics and the Structure of Industry'. The pairs of texts are from 15 of the 30 disciplines in the BAWE corpus, and from 9 of the 13 genre families. As the shaded area in Table 1 suggests, the composition of the corpus thus reflects the genre families that are more frequent in the BAWE corpus (i.e. Essays and Methodology Recounts) as well as the disciplines of courses that are popular with Chinese students and where an English match was found: the majority of texts are from Engineering, Food Sciences and Biology (20–32 each); other texts are from Business, Law, HTLM, Linguistics, and Sociology (10–16 each); and a small number are from Economics, Agriculture, Politics, Cybernetics, Psychology and Publishing (2–6 each).

Further details about these disciplines and genre family categories can be found in Gardner and Nesi (2013).

The 374,835 word Han CH-EN corpus (Table 2) comprises 156 texts (78 pairs), with an average length of 2403 words. Typically where assignment deadlines are set, for example at 3000 words, a margin of 10% is set so that students can write

Table 1
Genre Family and Discipline of texts in the Han-CH-EN corpus.

	ENGN	FDSC	BIOL	BUS	LAW	SOC	HTLM	LING	ECON	POL	AGR	PUB	CYB	PSY	TOTAL
Essay	2	2	0	10	10	8	4	6	4	4	2	0	2	0	54
Methodology Recount	12	22	10	0	0	0	0	0	0	0	0	0	0	0	44
Critique	2	0	2	0	2	2	2	4	2	0	0	0	0	2	18
Case Study	4	0	0	6	0	0	4	0	0	0	0	0	0	0	14
Explanation	2	0	6	0	0	0	0	0	0	0	2	0	0	0	10
Design Specification	8	0	0	0	0	0	0	0	0	0	0	0	0	0	8
Proposal	0	0	2	0	0	0	0	0	0	0	0	2	0	0	4
Literature Survey	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2
Problem Question	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2
TOTAL	32	26	20	16	12	10	10	10	6	4	4	2	2	2	156

Table 2
Words and Texts in the Han CH-EN corpus.

	Texts	Words	Average Text Length
Chinese (CH)	78	170,227	2182
English (EN)	78	204,608	2623
Total	156	374,835	2403

between 2700 and 3300 words. Our data shows that English students tend to write more than Chinese students, although all texts were of an acceptable length.

4.2. Identifying transitions

Unlike BAWE, the Han CH-EN corpus is tagged with the Penn Treebank tagset (Marcus et al., 1993). To identify transitions in the corpus, we started from a query in SketchEngine, [tag = "SENT"] [tag = "RB.*|CC|IN"], to identify all adverbs, conjunctions and prepositions in initial sentence position. This was repeated for items following a semi-colon³. Searches for items identified as transitions in the literature were also conducted.

We essentially follow Hyland's definition of 'transitions':

Transition markers are mainly conjunctions and adverbial phrases which help readers interpret pragmatic connections between steps in an argument. They signal additive, causative and contrastive relations in the writer's thinking, expressing relationships between stretches of discourse. It is unimportant whether items here contribute to syntactic coordination or subordination but to count as metadiscourse they must perform a role internal to the discourse rather than the outside world, helping the reader interpret links between ideas. (Hyland, 2005:50).

Concordance lines were examined manually to determine if indeed the items identified were functioning as transitions. Examples such as *however* in *however they please* (CH3ESLAW-0410b⁴) were thus excluded.

A manual examination was also needed to identify only transitions performing a role internal rather than external to the discourse. This was not always straightforward, but the internal-external distinction is key to understanding the concept of transitions and identifying them for the study.

The question also arises on whether intra-sentential connective items should be included as metadiscourse resources. Studies such as Cao and Hu (2014) and Gardezi and Nesi (2009) that focus on cohesion tend to follow Halliday and Hasan (1976)'s categories of Additive, Adversative, Temporal and Causal Cohesion, and to exclude connectors such as *because*, *although* and *since* when they are used intra-sententially. In contrast, we follow Hyland (2005) and Martin and Rose (2003/2007) who both recognize that intra-sentential connectors can serve the functions of internal transitions. This also allows us to include cases of clauses joined by semi-colons and by transitions, which would otherwise be borderline.

In the present study we start with corpus-based frequency counts in order to focus our attention on areas of similarity and difference between Chinese and English writers, and then move to a more nuanced interpretation of specific instances of a transition in context.

4.3. Quantifying transitions

The following procedures were used to produce the descriptive statistics for the transitions identified:

- A data base of transitions by text was created in Excel. This included information about the position of the transition in the sentence. In this paper we focus on Sentence Initial (SIP) vs Non-Sentence Initial Positions (Non-SIP), with further detail in specific examples.
- Using SPSS, we calculated the *observed absolute frequencies* (N) and *observed relative frequencies* (per 1000 words) so that comparisons could be made across groups of different sizes. Significant differences were calculated using independent-samples t-tests in SPSS.
- Outliers were identified using SPSS. The inclusion of outliers made no difference to the statistical significance of totals or groups. So to avoid idiosyncratic influences, they have been omitted from the analysis here.
- Where there are five or fewer text pairs in a cell (e.g. a discipline or genre family), we have assumed this is too small and potentially idiosyncratic for trustworthy generalization and focused our analysis on cells with greater representation to aim for more robust findings.

These procedures are discussed in greater detail in Han, 2018.

5. Findings

This section presents an overview of the transitions identified and their frequency in the Han CH-EN corpus (5.1), followed by a focus on the transition *however* in terms of genre (5.2), discipline (5.3) and syntactic positions (5.4).

5.1. Transitions in the Han CH-EN corpus

In total, 46 transitions were identified in the Han CH-EN corpus. This includes many listed in the appendices of Hyland (2005) and others such as *meanwhile* and *whilst* uncovered through the corpus query language.

Although the absolute frequency of the 46 transitions in Han CH (Chinese writers) is lower than in Han EN (English writers) (1746 vs. 1870), the relative frequency (per 1000 words) is higher (10.500 vs. 9.171) (see Table 3). This difference is not statistically significant ($p > 0.05$).

This finding of no significant difference for overall frequency is similar to Granger and Tyson's for advanced French and English essays (1996:21). The relative frequency of around 10 per thousand words is closer to Hyland's finding for research articles (12.3) than for textbooks (28.1).

Table 4 shows the results for the 46 transitions identified, with their absolute frequency (N), their relative frequency per thousand words (ptw), and the P-value, with significant differences ($p < 0.05$) highlighted in bold.

Both Chinese and English writers used 44 of the 46 transitions, and no significant difference was found for the majority (34) of transitions. The frequencies for both English and Chinese writers, at 0.9 for *but* and 0.8 for *because*, suggest remarkable similarity in terms of frequency. *Therefore* was used more by English students, but the difference was not significant. Then the frequency of *so* (0.7ptw) and *although* (0.6ptw) were also remarkably similar. Together the first six transitions accounted for more than half of the transition use. The other 40 occurred less frequently.

The items in bold are transitions where significant differences were found. Chinese students used nine transitions (*since*, *thus*, *while* [comparison], *and*, *on the other hand*, *in addition*, *whereas*, *nevertheless*, *in contrast*) significantly more than English students. This suggests that, contrary to the general finding that Chinese students use a narrower range of transitions, successful Chinese university students were generally using a greater variety of transitions than their English counterparts.

Three transitions were used significantly more by English writers (*however* and *whilst* for consequence and comparison). Because of the similarity between *while* and *whilst*, these are both considered in a paper on transitions of contrast used significantly more by Chinese writers (Gardner & Han, 2018). Here we focus on *however* because it was the most frequent transition for both groups of writers, and because it is unusual in being used significantly more frequently by English writers than Chinese writers. Our investigation of *however* has two main aims:

To find out whether *however* is used more by English or Chinese writers, we ask:

1. Are there any patterns of use of *however* that can be identified for English writers or for Chinese writers? (4.2)

To find out whether *however* is used more in technical scientific discourse or discursive soft discourse, we ask:

2. Is *however* used significantly more frequently in any specific genres (4.3) or disciplines? (4.4)

5.2. Patterns of use of *however* in CH and EN

The investigation of patterns of use of the transition *however* starts with sentence position. The occurrence of *however* in Sentence Initial Position (SIP) and Non-SIP for L1 Chinese and L1 English student texts are shown in Table 5. Chinese and English students both used considerably more *however* in sentence initial position than in Non-SIP positions: The relative frequencies in SIP are 1.236 and 1.108 per 1000 words, compared to Non-SIP of less than 1 (0.204 and 0.803). There is no statistically significant difference for the relative frequencies of SIP *however* ($p > 0.05$), while Chinese students used significantly less Non-SIP *however* than their English counterparts ($p < 0.05$).

5.2.1. SIP and PIP *however*

Table 5 suggests there is no significant difference in frequency of *however* in sentence initial position:

Table 3
Transitions used by Chinese and English student writers.

	Chinese	English	P-value
Total (n)	1746	1870	
RelFreq (ptw)	10.500	9.171	$p = 0.070$

Table 4

Frequencies of transitions in the Han CH-EN corpus.

TM	Chinese		English		P-value
	N	ptw	N	ptw	
however	244	1.440	364	1.911*	0.047
but	188	0.937	189	0.907	0.873
because	138	0.844	159	0.888	0.813
therefore	132	0.842	211	1.049	0.318
so	88	0.789	130	0.740	0.783
although	106	0.636	127	0.613	0.871
since	109	0.628*	78	0.331	0.048
thus	92	0.601*	91	0.340	0.023
while (comparison)	99	0.553*	27	0.130	0.000
and	48	0.384*	4	0.027	0.032
furthermore	50	0.340	65	0.257	0.343
also	32	0.207	21	0.097	0.064
on the other hand	38	0.206*	9	0.037	0.000
in addition	29	0.202*	14	0.047	0.005
though	36	0.201	16	0.105	0.253
whereas	26	0.189*	14	0.069	0.024
nevertheless	35	0.179*	10	0.045	0.007
hence	32	0.166	46	0.238	0.347
moreover	24	0.123	16	0.056	0.095
while (consequence)	23	0.104	16	0.062	0.352
as a result	19	0.101	13	0.049	0.253
consequently	14	0.101	12	0.109	0.907
yet	20	0.077	35	0.160	0.115
in contrast	11	0.070*	3	0.012	0.025
besides	18	0.070	3	0.017	0.138
additionally	11	0.057	4	0.022	0.247
similarly	9	0.054	12	0.058	0.910
meanwhile	6	0.045	0	0.000	0.100
on the contrary	10	0.043	2	0.017	0.311
again	6	0.041	21	0.085	0.180
likewise	6	0.031	1	0.004	0.062
rather	8	0.028	7	0.034	0.737
at the same time	5	0.028	3	0.008	0.298
even if	6	0.027	15	0.072	0.115
even though	5	0.025	13	0.064	0.111
nonetheless	4	0.025	1	0.005	0.144
conversely	4	0.022	6	0.035	0.490
as a consequence	2	0.018	1	0.003	0.286
thereby	2	0.015	1	0.003	0.347
accordingly	2	0.014	7	0.058	0.243
by contrast	3	0.012	1	0.007	0.684
alternatively	2	0.001	6	0.029	0.360
still	1	0.007	4	0.016	0.393
whilst (consequence)	1	0.005	69	0.279*	0.000
correspondingly	1	0.004	0	0.000	0.320
further	1	0.004	5	0.024	0.187
whilst (comparison)	0	0.000	17	0.075*	0.001
in the same way	0	0.000	1	0.002	0.320

*indicates a significantly greater value ($p < 0.05$).**Table 5***However* in different positions.

Position	Chinese		English		P-value
	N	ptw	N	ptw	
SIP	204	1.236	219	1.108	0.491
Non-SIP	40	0.204	145	0.803*	0.000
Total	244	1.440	364	1.911*	0.047

1. When milk is beaten, an extremely delicate foam forms. *However*, this foam collapses almost immediately after it is formed. (CH2MRFS-6008d⁴)
2. It can also retard oxidation, and improve storage life, which are all very desirable features. *However* it is very important that this is a controlled process as too much SO₂ can give food an unpleasant smell. (EN2MRFS-6004d)

There is, however, a significant difference in paragraph-initial position ($p < 0.05$), with Chinese using it more paragraph initially (0.255 vs. 0.065 per 1000 words), as in this example:

3. ... Without a mechanism to effect a just allocation of responsibility, this seems to assert considerable hardship on the part of trustees.

However, New Zealand and Canadian authorities seem to point to the contrary. (CH3ESLAW-0410a).

Moreover, the distribution of *however* in PIP is greater for Chinese writers both in terms of the proportion of texts in which it occurs (30% vs 14%) and the proportion of students who use it (38% vs 20%).

5.2.2. Non-SIP *however*

L1 Chinese students used significantly fewer occurrences of *however* in Non-SIP positions than their L1 English counterparts. When taking a closer look at the Non-SIP *however*, we found there were differences in three main types of Non-SIP *however*: 1) Non-SIP *however* used in the middle of a clause; 2) Non-SIP *however* following a semicolon between clauses; 3) Non-SIP *however* used as a conjunction.

These will now be discussed in turn, with reference to the distribution data presented here in [Tables 6 and 7](#).

[Table 6](#) shows Chinese writers used significantly less clause medial *however* and significantly less *however* as a conjunction than English writers ($p < 0.05$), with no significant difference in frequency after a semi-colon. The information in [Table 7](#) allows us to check if the data is well distributed.

When we examine the distribution of clause medial *however* ([Table 7](#)), we see a consistent pattern of much greater clause-medial use in Han-EN by number and percentage of occurrences, texts and student writers. This is evidence of an established pattern: In terms of non-SIP, the majority of EN students (60%) use *however* clause medially most of the time (61.70%).

As a transition, *however* makes a contrast with what has gone before. When it occurs in a clause medial position, therefore, it is generally near the beginning of the clause, as in these extracts from postgraduate Sociology Essays:

4. Such a reading, *however*, does not exhaust the rich implication of Marx's theory. (CH4ESSOC-0319a)
 5. Chronic illness, *however*, may bring these distanced ideas into reality for the sufferer, creating what Giddens has termed a 'critical situation', or a major disruptive experience. (EN4ESSOC-0405a)

The use of commas is variable. This first year Engineering Case Study omits them:

6. They *however* do have a high level of control over the price of replacement parts and services. (CH1CSENG-0008a)

In clause medial positions, *however* typically follows the first element of the clause, which may be Subject, as in 4, 5 and 6, or Adjunct, as in this third year Cybernetics Essay:

7. In some areas, *however*, the opposite is true. (EN3ESCYB-6101)

The second case of Non-SIP *however* is that following a semicolon between clauses as in these postgraduate Psychology Critiques:

8. It is said they wanted to understand if the mood state was related to gamble; *however*, researchers only assessed the anxiety/arousal levels. (CH4CRPSY-0104b)
 9. The additional control experiment was essentially the same as the first experiment; *however* participants were also given a multiple-choice questionnaire as a post-test to determine how closely the participant had attended to the material. (EN4CRPSY-0171b)

Here, as in sentence initial position, *however* is used to highlight a contrast between two clauses. The semi-colon serves to bring the two clauses closer together into one sentence, although its function has been given as 'to separate two main clauses' ([Lea, 2014: R25](#)).

Table 6

Frequency of *however* in non-SIP.

	Chinese		English		P-value
	N	ptw	N	ptw	
Clause-medial	12	0.05	87	0.49*	0.000
After semi-colon	21	0.13	12	0.10	0.643
As a conjunction	7	0.04	41	0.20*	0.003

Table 7
Distribution of *however* in non-SIP.

	clause medial		after semicolon		as a conjunction	
	Chinese	English	Chinese	English	Chinese	English
Occurrences	12	87	21	12	7	41
Total	40	141	40	141	40	141
Percent	30.00%	61.70%	52.50%	8.50%	17.50%	29.10%
Texts	10	36	9	9	6	23
Total	78	78	78	78	78	78
Percent	12.80%	46.20%	11.50%	11.50%	7.70%	29.50%
Students	9	30	5	8	6	18
Total	32	50	32	50	32	50
Percent	28.10%	60.00%	15.60%	16.00%	18.80%	36.00%

In this case, although there is difference in absolute frequency between Chinese and English students' writing (21 vs. 12), the relative frequencies are close, both around 0.1 per 1000 words (see Table 6). No statistically significant difference was found ($p > 0.05$). This type of *however* accounts for more than a half (52.5%) those in Chinese students' writing, while less than 10% in English students', as Table 7 shows. Its distribution in texts and students is almost the same, both occurring in 9 texts and in around 16% percent of students.

Thirdly, students in both the Han CH and EN subcorpora also seem to use *however* as a conjunction to join two clauses into one sentence. Here are two examples:

10. The ones with 0.5 and 0.8 dispersal rate maintained slightly above the initial population size, *however* this could not be recognized strikingly when dispersal rate was 0.2. (CH2MRBIO-0036c)
11. The strength of the jury lies in trial by one's peers, thought to be the basic symbol of justice, *however* increasingly doubts have been raised as to the competency of a jury. (EN3ESLAW-0411a)

This suggests a lack of adherence by both Chinese and English student writers to the traditional grammar rules where only a semi-colon or a conjunction such as *but* can join two independent clauses into one sentence.

However as a conjunction occurs statistically significantly more in L1 English students' texts ($p < 0.05$) (Table 6), and is found across 23 texts written by 18 different English students (Table 7).

To sum up, patterns of non-SIP use are emerging where EN writers use *however* significantly more in the middle of clauses and between clauses functioning as a conjunction. Chinese writers use *however* more often following a semicolon, but no significant difference was found.

We now turn to the distribution of *however* across genres and disciplines in the Han CH-EN corpus.

5.3. *However* across genres

Table 8 shows the frequency of *however* in the four genres with more than five pairs of texts. Chinese students used *however* less than English students in Essays, Methodology Recounts and Case studies, and slightly more in Critiques. There is however no significant difference between the relative frequencies of Chinese and English students in any of these genres ($p > 0.05$). A comparison across genres shows that *however* is used more in Essays and Critiques than in Methodology Recounts and Case Studies.

5.4. *However* across disciplines

Table 9 shows the frequency of *however* in the five disciplines with more than five pairs of texts. Generally, English students have higher frequencies than their Chinese counterparts in both absolute and relative frequencies. A comparison across disciplines shows that for both L1 Chinese and L1 English students, writing in Law employs the densest *however* with a relative frequency of over 3 per 1000 words (3.127, 3.466), compared with Business of around 2 (1.754, 2.022), which is

Table 8
Transition *however* in four genres with more than five pairs of texts.

	Chinese		English		P-value
	N	ptw	N	ptw	
Essay	131	1.737	174	2.227	0.183
Critique	42	2.679	51	2.183	0.608
Methodology Recount	27	0.928	56	1.728	0.104
Case Study	26	1.706	32	2.012	0.672

Table 9Transition *however* in five disciplines with more than five pairs of texts.

	Chinese		English		P-value
	N	ptw	N	ptw	
Law	47	3.127	52	3.466	0.751
Business	36	1.754	44	2.022	0.591
Food Science	21	1.234	23	1.521	0.673
Biology	20	0.925	29	1.556	0.183
Engineering	13	0.490	71	1.651*	0.014

followed by the disciplines of Food Science, Biology and Engineering. This indicates that the relative distribution across disciplines is similar for CH and EN, and it is only in Engineering that statistically significant differences are found ($p < 0.05$).

In order to explore this further, we investigate the distribution of genres in Engineering.

As Table 10 shows, most of the text pairs in Engineering belong to the Methodology Recount and Design Specification genres, and here English writers use *however* substantially more in number (N) and in frequency per 1000 words.

Taken together then, Tables 8 and 9 suggest that there are similarities in the frequency of *however* by Chinese and English student writers across most disciplines and across most genres, where *however* is most frequent in Law and in Essays. Statistically significant differences are found in Engineering ($p < 0.05$), where clear distinctions emerge in Methodology Recounts and Design Specifications between Chinese and English writers.

6. Discussion

We now discuss our findings in relation to the questions about whether there are differences in use between Chinese and English writers in the Han CH-EN corpus and where transitions are frequent.

6.1. CH vs EN patterns of use

Our investigation of the different positions for *however* revealed three findings worthy of discussion.

First, the most usual position for *however* in both Han CH and Han EN subcorpora is at the beginning of a sentence (83.6% and 60.2% respectively). In other words, not only L1 Chinese students but also L1 English students tended to use *however* in the sentence initial position (SIP). This finding is important because previous studies (Granger & Tyson, 1996; Lee & Chen, 2009; Leedham, 2015) give the impression that it is only Chinese students who favour initial position.

In contrast, in clause medial position our findings were similar to others. We found that the number and percent of clause medial *however*s in the Chinese student writing is significantly less than that in the English student writing.

When *however* is used clause medially, it typically follows the Subject or the Adjunct. Using it in this way may appear more sophisticated, and it is therefore worth trying to understand when this works well. Here the concept of information flow with particular reference to given and new information proves useful. Consider example 12:

12. The monetarist approach *however*, emphasises the role of expectations in distinguishing the long-run Phillips curve and the Short-run Phillips curve (EN2ESECO-0399b).

Here the subject, *the monetarist approach*, has been introduced earlier in the text, and is presented as given or old information; this allows *however* to introduce the rest of the sentence. In this case *however* plays its role of introducing new information that contrasts with what was said earlier. This provides a useful rule of thumb for when to use sentence initial and when to use sentence medial *however*: sentence medial *however* can be used when the Subject is given information.

In other cases more semantically empty units appear at the beginning of a sentence, and here *however* adds to the rhythm and accentuates the end weight of information:

Table 10Transition *however* in genres across Engineering.

Engineering Text Pairs	Genres	Chinese <i>however</i> N (ptw)	English <i>however</i> N (ptw)
6	Methodology Recount	1 (0.090)	24 (1.794)
4	Design Specification	2 (0.566)	21 (1.173)
2	Case Study	4 (0.966)	10 (3.058)
1	Critique	2 (0.213)	8 (1.705)
1	Problem Question	1 (0.945)	7 (2.059)
1	Explanation	2 (1.765)	1 (1.079)
1	Essay	1 (0.523)	0 (0.000)
16	TOTAL	13	71

13. In some areas, however, the opposite is true (EN3ESCYB-6101c).
14. There are, however, many questionable areas in Herzberg's two-factor theory (CH1ESBUS-0271c).

Arguably, where students can effectively use clause-medial *however*, their writing will appear more fluent and convincing. We have not investigated this explicitly ourselves, but it is our impression, and other research has shown that clause-medial *however* is statistically more frequent than clause-initial *however* in published British and American academic writing (Yin, 2014:163).

Finally, the use of *however* as if it were a conjunction is worthy of comment. This is a notable feature of English student writing, and it reflects a trend that has been widely observed by teachers of English, as these excerpts from a discussion group⁵ in 2013 suggest:

Is 'however' a conjunction?

I'm not a hardline prescriptivist, but I do retain a few of the 'rules' that I continue to 'correct' when students flout them in their academic writing. I still tell them that 'however' is not a conjunction and that 'but' is often 'better'

Shall I just give up and start telling students that 'however' is indeed an alternative to 'but'? [A]

This message immediately provoked 15 responses, including:

The fact that we can place 'however' in the positions permitted to conjunctive adverbs and that we can't do the same with 'but' suggests that we have to maintain the distinction with students of grammar even if the punctuation issue may be a lost cause. [B]

Students I've spoken to get told at school not to use 'but' so they substitute 'however' assuming they can use it in the same way to link sentences. I think you owe it to your students to alert them to the negative impact this makes on an audience. [C]

I suspect we are fighting a losing battle against a tide of linguistic evolution, but, until the tide is strong enough to remove the negative impression that [C] mentioned, I think anyone involved in raising students' awareness of language should keep up the fight. [D]

The responses suggest that using *however* as a conjunction is not acceptable in academic writing, and that we owe it to our students to explain why. While many Chinese students who have learned English grammar systematically will readily grasp and accept that *however* is not a conjunction, the task of explaining this to English students may feel at times, as D says 'like a losing battle'.

6.2. Transition *however* in genres and disciplines

However is used more often in Essays and Critiques (Table 8) and in Law (Table 9). This supports Hyland's observations about the use of such metadiscourse to develop arguments in the soft sciences, as this example from a third year Law Essay weighing up pros and cons of an argument suggests:

15. The case of Harold Shipman illustrates the possibility of individuals seeking to control the mortality of their patients, and their potential to do so would be fostered by a regime such as the Netherlands'. *However* the current system is not impervious to abuse as the case of Shipman illustrates. Legalising PAS may even improve the system by introducing a regulatory body to control and detect abuses. (EN3ESLAW-0119c)

Here *however* contributes to the weighing up of the two sides in the argument, which is further evidenced through the tentative nature of claims on both sides (*possibility, seeking to, potential, may even...*). It is expected that the Essay will end with the student developing a particular point of view, based on the arguments.

There is also evidence of use in Engineering Methodology Recounts and Design Specifications that supports Biber's claims that *however* is used in such scientific writing, as in this example from Engineering:

16. Hence roughly 1.5m should have been needed for a parabolic velocity profile to develop. *However*, looking at Graph 1, we see that the flow has fully developed, i.e. there are no more adverse pressure gradients, from roughly 0.5m down the pipe. Therefore this result is flawed. (EN2MRENG-0023d)

The arguments here are much more logical and based on more factual observations. As such, all students who conduct the same experiment are expected to reach the same conclusions.

In both of these examples, *however* does more than simply contrast; it also counters the argument in the first sentence – the function identified by Martin and Rose (2007). In fact, *however* in the Chinese data regularly occurs with *there* and *it* used non-anaphorically, so simply to contrast:

17. In terms of structural and institutional changes, the Dutch economy did indeed shift from the agriculture sector to the service sector, but they did not show a change from personal to large-scale firms. *However*, there was a high degree of urbanization and religious toleration, and the state played the role of securing property rights of merchants and entrepreneurs. (CH1ESECO-0071a)

In this Chinese example, the reader has to work hard to make connections between the information on either side of the *however*. In the English examples, *however* is often followed by *this* which introduces a clause that makes the contrast explicit (*this is not the case*). The additional inclusion of *not* is also effective in highlighting the contrast being made:

18. Because the flow has been tripped to turbulent in the turbulent flow, one would assume that its Reynolds number is less than that for laminar flow. *However this is not* the case as the apparatus is in a transitional stage and therefore the Reynolds numbers should be similar for both procedures. (EN2MRENG-0023d)

This clause has an anaphoric effect, referring the reader back with the cohesive *this* as well as reinforcing the contrast.

We have thus found evidence both statistical and discursive to support the importance of transitions such as *however* in building arguments in both the soft persuasive disciplines and the hard experimental sciences. We have also considered the company transitions keep, and how they build arguments in different ways.

7. Conclusions

Because the Han CH-EN corpus was developed by matching texts by Chinese writers with those by English writers, like Chinese students, it does not span the whole range of disciplines and genres of the BAWE corpus. Nevertheless it was broad enough to enable us to answer questions about where transitions are frequent. In this paper we have focused on the form and position of transitions. Thus while we can say that *however* is used more, for instance, by English writers in Engineering, it may well be that the same adversative meaning is expressed in different ways by Chinese writers. This is an area for future research.

Yin's (2014) study demonstrates that the choice and frequency of transitions varies across registers (e.g. across academic writing, news writing, lectures and conversation). It is perhaps therefore not surprising that earlier studies that compared Chinese student writing with published research in the same discipline (e.g. Lee & Chen, 2009; Lei, 2012) also found differences. Lei (2012) found that Chinese students used transitions more frequently, but a narrower range of transitions, and notably used *however* less frequently than published authors.

In contrast, in our corpus which matches student texts by discipline and genre, we found no significant difference in the overall number of transitions. This adds weight to a growing body of studies that seeks and does not find significant differences in Chinese vs English university student writing. For example, Geng and Wharton's (2016) study of evaluative language:

When we began this research, our underlying assumption was that we would discover that L1 Chinese and L1 English writers were using at least some aspects of evaluative language differently. Our finding that this was not the case challenges 'deficit' assumptions, and lends support to those who argue that the challenges of academic literacy, and of specific genre proficiency, are not a matter of language proficiency or language experience per se. (p.89).

7.1. Patterns of use by Chinese and English writers

The findings on sentence position were in some respects predictable from the literature in that Chinese students do use *however* less flexibly in clause medial positions, where we were able to probe the nature of discursive contrasts and suggest rules of use. In position, however, there is a twist. Chinese writers conformed more to traditional grammar expectation in terms of their use of commas surrounding *however*, their use of *however* following semi-colons, and their lesser use of *however* as a conjunction. In contrast, English students were found to overuse *however* as a conjunction. While the reasons for this were not investigated, the discussion suggests that this is still a use worth addressing. Moreover, a recent study of how transitions in general are taught in 39 EAP textbooks (Walková, 2020) suggests there is much for teachers to do to supplement textbooks in this area. Our findings related to when medial *however* is appropriate in particular with reference to given and new information (5.1) and related to *however* being important in both discursive and scientific discourses could usefully inform EAP teaching materials.

Among the transitions identified in the Han CH-EN corpus, the most frequent was *however*, which was also used significantly more by English writers, and became the focus of this paper. Other transitions identified as being overused (*besides*) or underused (*additionally*, *furthermore*) by Chinese writers in the literature (Lee & Chen, 2009) showed no significant differences. Our study supports Leedham's (2012) finding that Chinese writers in BAWE overuse *on the other hand*, and extend it by providing a list of 46 transitions with comparative data.

7.2. Differences in genres and disciplines

Using the closely matched Han CH-EN corpus to examine *however*, we discovered that significant differences between Chinese and English writers in the corpus as a whole do not extend equally to all disciplines and genres. Most strikingly, perhaps, we found many disciplines and genres where there were no significant differences observed between successful Chinese and English writers, but where there were clear parallel differences between disciplines (e.g. Law uses *however* more than Business which uses it more than Biology) and between genres (e.g. Critiques uses *however* more than Case Studies). This suggests that both Chinese and English students are being successfully acculturated into the linguistic conventions of their disciplines.

We did find substantial differences in Essays and Critiques, and statistically significant differences in Engineering, however, and in doing so we have found evidence to support both Biber's claim that *however* is used frequently in technical writing (in our case in Engineering) and Hyland's claim that *however* is used more in discursive writing (in our case in Essays and Critiques). As most of the writing in Engineering was not in Essays, these represent two distinct findings.

This paper has demonstrated the value of a closely matched corpus. The Han CH-EN corpus data is limited by the number of texts submitted to the BAWE corpus by Chinese students not educated in the UK prior to university. Future research could usefully extend our findings by developing a similar closely-matched corpus but with a greater range of disciplines and genres or for students with different first languages.

This paper has also shown the importance of probing contexts of use rather than frequency data alone. English writers use *however* more often than Chinese writers in sentence medial positions, but this does not mean that Chinese writers 'underuse' it, or that English writers 'overuse' it (both of which have negative connotations of deviant behavior). We find that English writers use it appropriately in clause medial positions to mark contrasts in information flow; they use it sparingly following a semicolon, where they could use it more; and they may use it inappropriately to function as a conjunction, where they could use it less. In other words, even successful academic writing by native speakers may contain features to which some readers will object.

Endnotes

1. We are using the term 'genre' here to include published research articles and textbooks, as well as genre families of student writing such as essays and case studies.
2. Although Biber's conception of linking adverbials is more developed in terms of types of adverbial than Hyland's concept of transition markers which are explored as part of metadiscourse, the studies reported here are all based on corpus searches, so the findings can be compared. We appreciate the comment from an anonymous reviewer who notes that the concept of 'linking adverbials' is more complex than we suggest here and acknowledge that we focus on form and position rather than meaning.
3. *And*, exceptionally, was not considered following a semi-colon.
4. The unique code for each text incorporates L1 (Chinese/English), Level of Study, Genre Family, Discipline, and the ID number for the student and text. So CH2MRFS-6008d is a Chinese (CH) 2 nd year (2) Methodology Recount (MR) Food Sciences (FS) text d written by student 6008.
5. BAALMAIL Discussion List October 23, 2013; names of contributors withheld.

Statement

Chao Han built the Han CH-EN corpus and analysed the data under the supervision of Sheena Gardner. Sheena Gardner wrote the paper with input from Chao Han.

Acknowledgements

The data for this paper draws on the BAWE corpus (www.coventry.ac.uk/bawe). The British Academic Written English (BAWE) corpus was developed at the Universities of Warwick, Reading and Oxford Brookes under the directorship of Hilary Nesi and Sheena Gardner (formerly of the Centre for Applied Linguistics [previously called CELTE], Warwick), Paul Thompson (formerly of the Department of Applied Linguistics, Reading) and Paul Wickens (Westminster Institute of Education, Oxford Brookes), with funding from the ESRC (RES-000-23-0800).

The paper is also supported by the Fundamental Research Funds for the Central Universities of China Project number: 2020jbkzy009 awarded to Chao Han.

References

- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam/Philadelphia: John Benjamins.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.
- Cao, F., & Hu, G. (2014). Interactive metadiscourse in research articles: A comparative study of paradigmatic and disciplinary influences. *Journal of Pragmatics*, 66, 15–31.
- Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language, Learning and Technology*, 14(2), 30–49.

- Gardezi, S. A., & Nesi, H. (2009). Variation in the writing of Economics students in Britain and Pakistan: The case of conjunctive ties. In M. Charles, D. Pecorari, & S. Hunston (Eds.), *Academic writing: At the interface of corpus and discourse* (pp. 236–250). London: Continuum.
- Gardner, S., & Han, C. (2018). Transitions of contrast in Chinese and English University student writing. *Educational Sciences: Theory & Practice*, 18(4), 861–882. <https://doi.org/10.12738/estp.2018.4.0067>
- Gardner, S., & Nesi, H. (2013). A classification of genre families in university student writing. *Applied Linguistics*, 34(1), 1–29.
- Geng, Y., & Wharton, S. (2016). Evaluative language in discussion sections of doctoral theses: Similarities and differences between L1 Chinese and L1 English writers. *Journal of English for Academic Purposes*, 22, 80–91.
- Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, 15(1), 17–27.
- Halliday, M., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Han, C. (2018). The use of English transition markers in Chinese and British university student writing. PhD Thesis submitted to Coventry University.
- Ho, V., & Li, C. (2018). The use of metadiscourse and persuasion: An analysis of first year university students' timed argumentative essays. *Journal of English for Academic Purposes*, 33, 53–68. <https://doi.org/10.1016/j.jeap.2018.02.001>
- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. London: Continuum.
- Kennedy, C., & Thorp, D. (2007). A corpus-based investigation of linguistic responses to an IELTS academic writing task. *Studies in Language Testing*, 19, 316–377.
- Knoch, U., Macqueen, S., & O'Hagen, S. (2014). An investigation of the effect of task type on the discourse produced by students at various score levels in the TOEFL iBT writing test. *ETS Research Report*. <https://doi.org/10.1002/ets2.12038>
- Lea, D. (Ed.). (2014). *Oxford learner's dictionary of academic English*. Oxford: Oxford University Press.
- Lee, D., & Chen, X. (2009). Making a bigger deal of smaller words: Function words and other key items in research writing by Chinese learners. *Journal of Second Language Writing*, 18(4), 281–296.
- Leedham, M. (2012). Combining intuition with corpus linguistic analysis: A study of marked lexical chunks in four Chinese students' undergraduate assignments. *Nordic Journal of English Studies*, 11(3), 155–187.
- Leedham, M. (2015). *Chinese students' writing in English: Implications from a corpus-driven study*. London/New York: Routledge.
- Lei, L. (2012). Linking adverbials in academic writing on applied linguistics by Chinese doctoral students. *Journal of English for Academic Purposes*, 11(3), 267–275.
- Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2), 313–330.
- Martin, J., & Rose, D. (2007/2003). *Working with discourse: Meaning beyond the clause* (2nd/1st). London: Continuum.
- Milton, J. (1999). Lexical thickets and electronic gateways: Making text accessible by novice writers. In C. N. Candlin, & K. Hyland (Eds.), *Writing: Texts, processes and practices*. London (pp. 221–243). London: Longman.
- Walková, M. (2020). Transition markers in EAP textbooks. *Journal of English for Academic Purposes*, 46, 100874. <https://doi.org/10.1016/j.jeap.2020.100874>
- Yin, Z. (2014). *Linking adverbials in English*. Unpublished PhD thesis. New Zealand: Victoria University of Wellington.

Chao HAN, MSc TESOL (Bristol), PhD (Coventry), is Lecturer in English at Lanzhou University. In his PhD supervised by Sheena Gardner and Hilary Nesi, he created the Han CH-EN corpus, a matched Chinese-English subset of the BAWE corpus, to investigate transition markers as metadiscourse in academic writing and EAP.

Sheena GARDNER is Professor of Applied Linguistics and member of the Centre for Arts, Memory and Communities at Coventry University. Her research in Educational Linguistics centres on academic genres and registers in the BAWE corpus of university student writing, feedback, and transfer of learning from EAP to Disciplinary contexts.