Permanent WRAP URL:
http://wrap.warwick.ac.uk/75219

A note on versions:
The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP URL' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications

Evaluative language in discussion sections of doctoral theses: similarities and differences between L1 Chinese and L1 English writers

Yifan Geng
School of International Studies
University of International Economics and Business
Beijing
China.

Sue Wharton (corresponding author)
Centre for Applied Linguistics
University of Warwick
Coventry
UK
s.m.wharton@warwick.ac.uk

Evaluative language in discussion sections of doctoral theses: similarities and differences between L1 Chinese and L1 English writers

Introduction

Previous research (e.g. Bitchener & Basturkmen, 2006, Chatterjee, 2008) has shown that the discussion section of the doctoral thesis is a challenging part-genre (Dudley Evans, 1998) to write. The doctoral thesis as a whole is a high-stakes genre, in which writers must negotiate a dual identity: simultaneously as a student presenting him/herself for examination, and as an expert demonstrating his/her credentials for membership of a disciplinary elite (Carter, 2011; Carter & Blumenstein, 2011; Koutsantoni, 2006; Paltridge & Starfield, 2007; Swales, 2004). The discussion section is arguably the part of the thesis in which these two identities come into conflict most acutely, since it is here that a writer is required to comment evaluatively on his/her own work and also set it in the context of other work in the same field, with the overall goal of showing the reader why the contribution of the doctoral research is significant.

A discussion section requires skilful deployment not only of citation for attribution, in order to demonstrate the writers' display of knowledge, but also of citation for evaluation, in order to demonstrate that the writer has succeed, to some extent, in transforming knowledge and creating new knowledge (Parkinson, 2011). To make the case for the significance of their own results, writers need to position themselves and negotiate their own voice in the context of utterances made by others in the research community. Bitchener & Basturkmen (2006) and Chatterjee (2008) both show how this task is cognitively and linguistically challenging to thesis writers who are, after all, attempting the genre for the first time. Bitchener & Basturkmen (2006) discuss supervisors' perceptions of where the difficulties lie; supervisors emphasise the need for students to make sufficient links to previous literature when interpreting their results, and to negotiate knowledge claims against the backdrop of published knowledge. Despite recognition of these difficulties there is, within the

1

body of research on the doctoral thesis as a genre, comparatively little focus on the deployment of evaluative language in the part-genre of the discussion.

Research on evaluative language in academic texts has tended, rather, to focus either on published research articles (e.g. Bruce, 2014) or on student writing at earlier stages of education. Within this second category, researchers have often conducted comparisons of argumentative discourse produced by L1 and L2 writers, and have tended to conclude that L2 student writers rely on a smaller range of epistemic expressions and grammatical resources to realise interpersonal positioning (Chen, 2010; Gabrielatos & McEnery, 2005; Hyland & Milton, 1997; McEnery & Kiffle, 2002).

Hyland and Milton (2007), comparing epistemic markers in argumentative essays by Hong Kong and UK undergraduates, found that the ten most frequently occurring markers accounted for three quarters of the total markers in the Hong Kong corpus. The Hong Kong corpus contained more markers of certainty, whereas the UK corpus contained more markers of probability and tentativeness. Similarly, Chen (2010) found that L1 Chinese student writers were more likely than L1 English writers to employ the item *sure*, but less likely to employ items such as *possible (-ly)* or *probably*.

McEnery & Kiffle (2002), studying the writing of Master students, found that L1 Eritrean writers relied on a narrower range of epistemic markers than did L1 English writers, tending to overuse modal verbs and adverbs of possibility. Gholami, Nejad, & Pour (2014), studying the writing of Iranian undergraduates, concluded that they used a low frequency of emphatics and attitude markers in conclusion sections of argumentative essays. Hinkel (2003) notes that a group of Asian student writers used notably more amplifiers and emphatics in their argumentative essays than did their L1 English counterparts.

The widely-held view that L2 students use a smaller repertoire of interpersonal language when producing argumentative writing seems, then to be based mainly on

research into the writing of students at pre-doctoral stages of education. For students in an English medium doctoral programme and writing a thesis in English, the situation is different: they are likely to be experienced in English academic writing whatever their L1 is. In Swales' (2004) terms, they are likely to be at least 'narrowly English proficient' and quite probably 'broadly English proficient'.

The current research investigates the use of evaluative language in a corpus of doctoral discussions written by L1 Chinese and L1 English writers. Its first research question, based on the corpus as a whole, is: What patterns of textual evaluation are salient in this part-genre? Its second research question, based on a comparison of the L1 Chinese and L1 English sub-corpora, is: are there any systematic differences in the patterns of textual evaluation which the two groups employ?

Research design

To answer these questions, we compiled a corpus consisting of twelve discussion sections of theses in the discipline of Applied Linguistics, all successfully examined in the same university department during a relatively short time span. The writers had all participated in research activities organised by the department (methods seminars, critiques of research papers etc.) but had not received specific training in doctoral thesis writing outside of the supervision process itself.

Six of the discussion sections were from theses by L1 Chinese writers and six from L1 English writers. The assignment of a writer as an L1 Chinese or L1 English speaker was based on a Chinese- or Anglophone- sounding name as well as educational and other general background information mostly found in the Acknowledgments of the theses. The resulting corpus contains 118, 971 words. The average word length per text is 9,914, with 4,849 words in the shortest text and 13, 620 in the longest (see Appendix for details of each text).

This is a small, specialised corpus designed to answer a specific research question (Hunston, 2002); as argued by McEnery, Tono & Xiao (2006), the size,

representativeness, balance and sampling of the corpus should all be determined by the research question it is used to address. A small corpus was appropriate for the 'qualitative, contextually informed analyses' (Flowerdew, 2004: 18) which formed the basis of our corpus annotation. As will be discussed below, texts in the corpus were annotated manually at clause level and across clauses, with instantiations of interpersonal meaning being highlighted and categorised via a qualitative, interpretive process. This process of analysis and annotation was labour intensive, requiring sensitive consideration of co-text and discussion between analysts. Therefore a relatively small corpus was essential. Illustrations of our interpretations of sections of text are given below as we explain our analytical framework and discuss our results.

Approach to analysis

As was argued above, the discussion section of a doctoral thesis relies heavily on evaluative language, since its purpose is to critically discuss the researcher's findings and how they fit into the wider field, making and sustaining arguments across sections of text (Bunton, 1999). A suitable framework for investigating evaluative language in such a context is Martin & White's (2005) Appraisal Theory, which examines 'what is at stake interpersonally both in individual utterances and as the text unfolds cumulatively' (White, 2001: 8) and which presents interpersonal meaning-making as primarily a matter of negotiating a position vis a vis prior utterances and the actual or potential responses of others in a given communicative context.

The Appraisal framework deals separately with language users' attitudinal positionings (via the Attitude sub-system) and positionings regarding propositional meanings (via the Engagement sub-system). The Attitude sub-system categorises expressions of attitude into Affect (expressions of emotion) Judgement (evaluations of human behaviour) and Appreciation (evaluations of things or phenomena). The Engagement sub-system categorises propositional statements on the basis of whether they acknowledge the possibility of dialogue or not (*heterogloss* or

4

mongloss) and then, if the possibility of dialogue is acknowledged, on the basis of whether space for dialogue is opened up (*expansive heterogloss*) or shut down (*contractive heterogloss*). A full description of the Appraisal framework is outside the scope of this article, and readers are referred to Martin & White (2005).

A number of previous studies have used the Appraisal framework to investigate the interpersonal dimension of academic language. Hood (2005) found that published writers and Hong Kong student writers demonstrated different preferences for Attitude options in the introductions of their respective texts; student writers tended to rely on Judgement (e.g. '[they] **neglect** to observe…') and Affect (e.g. 'they usually feel…**depressed**'), which constructed their texts as more personal and subjective than the published texts, which relied more on Appreciation (e.g. 'Peer review is a **useful** technique') (p. 27). Liu & Thompson (2009) contrasted Attitude choices in two essays, one in English and one in Chinese, written by the same L1 Chinese undergraduate student. They found fewer Judgement items in the Chinese text, and suggested that this reflected an emphasis on 'the use of language and rhetoric to achieve social harmony' (Liu & Thompson, 2009: 7).

Swain's (2010) analysis of Attitude and Engagement in high and low score discussion essays by L2 English undergraduates suggested that use of Engagement, rather than use of Attitude, was the factor which best differentiated the high from the low scoring group. The high score set contained a greater amount and wider range of Engagement resources, as well as a better balance between expansive and contractive options. Similar findings are reported by Brooke (2014) who notes that higher rated undergraduate EFL writers create argumentation by using more interpersonal resources for attribution and authorial endorsement or disendorsement. Wu (2007), who compared the deployment of Engagement resources in high and low rated essays by Singapore undergraduates, found a greater prevalence of expansive engagement in high rated essays and a greater prevalence of contractive options in low rated essays. Particularly, writers of lower rated essays seemed to rely on Proclaim-Pronounce (a category which explicitly marks the author's commitment to a proposition) and made less use of the categories of

Disclaim: Counter/Deny (two categories which present the authorial voice as rejecting a contrary proposition). In contrast, the higher rated essays made frequent use of these categories as a strategy to introduce alternatives in order to reject them.

It seems clear from previous research that the Engagement sub-system of the Appraisal framework is a fruitful basis for the investigation of evaluative language in discussions of doctoral theses. In the next section, we give an outline of this subsystem and explain how we annotated our corpus. In our presentations of results, we explain the categories of this sub-system in much more detail, while simultaneously discussing examples from our own coded data.

Analytical framework and text annotation

In outline, our framework is as follows:

Figure 1 Analytical framework used in this study, adapted from Martin & White (2005)

Readers already familiar with Martin & White (2005) will notice that our diagram includes an option, Justify-from-Data, which is not part of their original framework. This is a category which emerged from our corpus investigation, and we propose that it may be specific to the part-genre of discussion sections of doctoral theses in the field examined. This category, as well as other categories shown in the above figure, will be explained and exemplified with reference to our own coded data after we have presented our approach to text annotation.

Texts in our corpus were annotated according to the above categories using UAM Corpus Tool, a software programme which allows the user to define an annotation scheme and then apply categories from the scheme to user-selected stretches of language within the texts of the corpus. Text annotation using a functional

6

framework is, of course, an interpretive act which involves a number of decisions. The first of these was to decide whose interpersonal meanings we wished to code. Following our research questions, we were interested only in instances of language which expressed the writer's own evaluations and interpersonal meanings. Stretches of language which expressed the interpersonal positioning of any other voice – for example, that of cited literature – were outside the scope of our study and were not selected for coding.

A second decision involved the selection of text span for coding. Interpersonal meanings are rarely carried by individual lexical items, so for each identification of text realizing a category we highlighted that span of text which seemed to us to do the most work in realizing the interpersonal meaning. In this sense our methodology differed from that used by Hyland (2005, 2010) or Biber (2006) in their larger scale corpus-based studies of evaluation in academic writing. Hyland (2005, 2010) worked from an 'a priori' list of some 300 evaluative language items identified from previous research; Biber (2006) focused on three pre-determined categories of evaluative language (modal and semi-modal verbs, stance adverbs and stance complement clauses). In our own research, in contrast, we selected any and all stretches of language which expressed the writer's evaluations in context.

Being conscious of the interpretive nature of our selections and analyses, we discussed all stretches of language where we were in doubt about appropriate coding. The first analyst made notes on any problematic items in a separate word file before sending them to the second analyst for comment; points were then discussed in a meeting. The following extract from notes is an example. It concerns a doubt about how to interpret authorial position towards cited literature:

Figure 2 Extract from analyst notes

In the meeting which followed, both authors used UAM CorpusTool to retrieve this example and looked at co-text to see if they could observe the text writer's clear alignment with Benson (2001). The discussion helped us to realise that it would sometimes be necessary to look at long stretches of co-text before alignment or disalignment with a cited source could be discerned.

UAM corpus tool contains a number of features which helped us to take a principled approach to assigning interpersonal language features to categories in the Engagement system, and to maintain consistency in our coding. The results of annotation can be viewed in a variety of ways, for example all instances of a particular analytical category, or all instances of a particular lexical item which appear in any analytical category. Figures 1 to 3 in the Supplementary Data file, show some examples.

By viewing our results in different ways and discussing them, we were able to constantly revisit our coding decisions and alter our decisions regarding the coding of individual segments if we found inconsistencies. Through this recursive process of coding, checking and re-coding, including detailed discussion on points of doubt and decisions of principle, we were able to arrive at a robust and consistent series of annotations.

In the section below, we discuss patterns found in our corpus while simultaneously giving illustrations of how the various categories of evaluative meaning were realized. By giving examples of coded data, we seek to show our interpretations of the categories. By commenting qualitatively on how texts in our corpus 'typically' made use of these categories of evaluative language, we address our first research question.

8

Results and discussion

The corpus as a whole

The corpus of 12 texts yielded a total of 5300 occurrences of Engagement realisations, with the highest number in one text of 658, the lowest 183, and the average 442 (see Appendix for more details of each text). Overall, there was one occurrence of an engagement realisation every 22.4 words. This rate of occurrence is comparable with that found in Hyland's (2010) metadiscourse research on a much larger corpus, and is therefore a good indication that our smaller corpus was adequate in size for our research questions.

The broad distribution of Engagement resources in our corpus was as follows:

Figure 3 Engagement in the whole corpus

The first point of interest is the relatively high proportion of Monogloss (bare assertion) realisations. Monogloss does not construe a background of diverse viewpoints for the proposition being advanced; it construes a reader who holds the same viewpoint or who will accept the proposition advanced as unproblematic. Given the dialogic purpose of Discussion sections, the relatively high proportion of Monogloss is counter-intuitive. However, qualitative examination of instances indicates that most are metalanguage, such as a summary of a section, cross-references to results that have been presented, e.g.:

> *1.   This chapter brings together the data findings and explores Chinese students' conceptions relatable to learner autonomy theory in the literature.* (Text 2: 261)

Also frequent are statements of 'fact' about the writer's context which the reader would not be expected to dispute, e.g:

2.   As China is undergoing a rapid transition from a planned socialist economy to a market socialist economy, some values from the West are being incorporated into the Chinese traditional value systems and some notions of LC are also promoted in the recent educational reform. (Text 4: 238)

Less frequent, though interesting from a genre perspective, are propositions about research or teaching which the reader construes as unlikely to be disputed by the readership of the thesis, as in this example:

3   ***…the effects of exams on learner motivation depend on how teachers present them in the classroom***. (Text 5: 209).

Regarding Heterogloss, the number of contractive realisations, (n=2033), which acknowledge alternative voices but shut down the possibility of dialogue, is roughly 1.5 times more than that of expansive features (n=1317) which open up the possibility of dialogue. This is in contrast to what might have been expected given, for example, Becher's (1990) categorization of ELT/Applied Linguistics as a soft-applied discipline, in which knowledge is "qualitative and reiterative" (Becher, 1990, p. 335) and therefore tends to openly acknowledge other points of view.

The following figure gives a more detailed distribution of Heteroglossic sub-categories in our corpus:

Figure 4 Heterogloss Features in the Whole Corpus

Within contractive Heterogloss, the subcategory of **Disclaim** (n=1027) contains slightly more instances than that of **Proclaim** (n=904). **Disclaim** involves the specific resources of **Deny** and **Counter**. These options invoke prior utterances or alternative views – but unlike expansive Heterogloss, they introduce such positions in order to refute or replace them. According to Martin & White (2005: 120) Deny is often realized by negation, and Counter by 'conjunctions and connectives such as *although, however, yet, and but*'. In our corpus, Deny and Counter were typically

used together in order to allow a writer to make very nuanced claims as to the value of their findings, as in the example below:

4.   The findings in this area of 'reasons' are therefore **_not_** generalizable and require further investigation, **_although_** they are of some interest in shedding light on an unexplored area. (Text 1: 210)

The subcategory Proclaim involves the resources Concur and Pronounce. Concur is concerned with formulations such as *naturally, obviously,* whereby writers explicitly construe a position as generally shared (Martin & White, 2005). In our corpus, such realisations tend to reference positions which could be construed as shared within the specialist readership of the thesis, rather than more generally. As such, they are interesting indicators of writers' assessments of community consensus.

5.   This tension between the administration's edict to teachers to "follow the textbook" and teacher autonomy naturally affected the students in the classroom. (Text 8: 253)

Pronounce refers to formulations that explicitly present a position as convincing, thus refuting any challenging position. Such formulations include 'explicitly authorial interventions' (e.g. *I contend…, The facts of the matter are that…*) and 'intensifiers with clausal scope' (e.g. *really, indeed*) (Martin & White, 2005: 127). An example from this corpus is:

6.   The importance of the inclusion of learner perspectives in the teacher appraisal process **_certainly_** has relevance. (Text 9: 324)

By choosing this realization the writer emphasizes their commitment to the proposition while simultaneously indicating awareness that others in the community may hold a different view.

11

The option **Endorse** differs from **Concur** and **Pronounce** in that it engages with an external voice which is presented as highly credible. By endorsing the external utterance, the authorial voice is positioned as in alignment with it and as sharing some responsibility for it. Typical realizations of Endorse given in Martin & White (2005: 134) are formulations such as '*the report demonstrates/shows/proves that...*'. However, in our corpus, realisations of **endorse** did not tend to rely on the reporting verbs suggested in Martina & White. Rather, it was through a reading of co-text that we could establish that the authorial voice was clearly associated with the attributed voice, as shown by the following example:

> 7. ***Kyriacou (1987) argues that*** identifying incompetence has a negative impact on teachers who may want to use the feedback to discuss problems, and this is <u>interesting</u> when related to Hannah's perspective [5.3.2.3]. (Text 9: 316)

In example 7 the authorial voice indicated its support for Kyriacou's (1987) proposition by the positive evaluation *interesting* and thus the writer positioned him/herself as aligned with the attributed voice. In our corpus, the single main use of Endorse was to indicate authorial alignment with cited literature, as in the above example.

The option Justify from Data, which we have placed within the contractive subcategory, is perhaps the most interesting in that it was generated inductively from our corpus and so may be specific to the part-genre which we have investigated. The category captures a position adopted by thesis writers as they referred to their research data within the discussion sections. We found that reference to own research data (usually data from interviews, questionnaires, and teaching journals) in the discussion sections was an option used to close down dialogue about the proposition being advanced, with the research data being used as a basis to convince the reader. Our category has some elements in common with White's (2012) proposal of a new dialogic contractive option of 'justify' in the genre

of newspaper editorials. In the following example, the writer uses research data to support his/her preceding proposition:

> 8.   Students reported they were rather confused by the whole thing. ***As one student commented***: 'I write the hook [name of a specific writing technique] in the Mr. Sun's class, he said no, you can't write this on it, so I think maybe it's not a part of academic writing.' (Text 8: 254)

Turning now to a discussion of Expansive Heterogloss in our corpus, we note that Entertain occurred 1055 times and Attribute only 262 times. For our writers, then, Entertain was the preferred means of opening up dialogic space for other potentially diverse views. For Martin & White (2005: 105) typical realizations of Entertain contain modal auxiliaries, adjuncts, and attributes of possibility; this was also true in our corpus. Entertain was frequently used by writers as they made 'generalised' claims based on their own research, for example:

> 9.   A failure to recognise teachers' past achievements, experiences and challenges is ***likely*** to alienate them and/or reduce their commitment towards a new reform agenda. (Text 7)

Within the Attribute subsystem, it is interesting to note that in our corpus the ratio of Distance (n=37) to Acknowledge (n=225) is about 1:6. Distance is a dialogic expansive option which marks the authorial voice as disaligned from the attributed position. Examples of Distance in Martin & White (2005) revolved mainly around reporting verbs such as *claim*, however, in our data realizations of Distance were more varied and more dependent on context. In our corpus, both L1 Chinese and L1 English writers typically used Distance in very similar ways – to argue against a finding from previous research in order to promote their own research. In the following example from an L1 Chinese writer, the author discussed new findings about the use of CLT approach (Communicative Language Teaching) in Chinese universities.

> **10.   *As indicated in the literature review (see 2.3), a basic reason for the inefficiency of CLT in China has been summarized by Hu (2002) as being that***

ideas advocated by CLT such as 'learning-by-doing' and equality between teacher and learner are in contradiction with Chinese teaching and learning culture, deeply influenced by Confucianism. ***But the findings showed that*** CLT is considered by certain participants as being fundamentally harmonious with the essence of Confucianism. In addition, ***unlike what was indicated in the literature review that the constraints of CLT are mainly at cultural level, the findings suggested that*** the major constraints of CLT seem to be more at technical and ideological levels instead. (Text 6: 242)

The author first referred to previous researcher Hu (2002) who argued the contradiction between the principles of CLT and Chinese teaching and learning culture; however, the author explicitly showed her disalignment from Hu's (2002) view by use of Counter (**But**). She then promoted her own finding via Pronounce (**the findings showed that…**), which reduced the dialogic space for others questioning the authorial proposition. The author continued to refer to previous literature that attributed the constraints of CLT to Chinese culture, but did this in order to distance herself from this view, as signaled by the semantic contrast (**unlike**). She then presented the opposite view that the constraints of CLT are more at technical and ideological levels but construed this authorial view as "located in some individual subjectivity, in some individual assessment of likelihood or of the available evidence" (White, 2001, p.6) by means of Entertain (**the findings suggested that…**). In contrast to the smaller dialogic space for alternative views created by Pronounce as used in the first part of the example this use of Entertain expanded such space for potential disagreement.

In the corpus overall, uses of Distance to disalign from previous literature were more likely to be co-articulated with contractive resources (Pronounce and Endorse) than expansive resources (Entertain) when these writers positioned their own different findings or views against the previous ones from which they disaligned. This preference for contractive options seems to be reasonable based on their function

to close down the dialogic space by increasing the "interpersonal cost of any rejection/doubting" of authorial propositions in the ongoing communication (White, 2001, p. 5).

Acknowledge, like Distance, references an external voice; but whereas Distance positions the writer as disaligned with that external voice, Acknowledge positions the writer as neutral towards it. In the following example, again referring to previous literature, the writer references an external voice but his/her own position vis à vis that voice is not made explicit:

11.   However, except for ***Daoud and Celce-Murcia (1979), who propose a post-use evaluation using the same checklist as for pre-use evaluation***, others do not provide concrete suggestions for in-use and post-use evaluation. (Text 1: 199)

In our corpus, then, the options of Endorse, Attribute and Acknowledge were all used by writers to position themselves vis à vis cited literature. As was seen in Figure 7, Endorse is the most frequently used of the three, followed by Attribute and then Distance. This sparse use of Distance in the current corpus may imply a relatively low number of different findings between the 12 authors' own and other research; alternatively, it may imply that the writers are hesitant about explicitly criticizing published findings due to the potential risk of being challenged or refuted by the reader.

In the sections above, we have examined evaluative language in our data in some detail, using the Appraisal framework. In this way we have begun to build at least a partial picture of the discussion section of a doctoral thesis in ELT/Applied Linguistics as a part-genre.

Our (admittedly small) dataset shows that writers use more Monogloss than may have been supposed. This arises partly through the use of cross reference or the assertion of 'simple' facts, but it also includes the use of bare assertions to reference views that the authors judge likely to be shared within their community. In these cases, Monoglossic realisations can be interpreted as functioning as invoked evaluation.

When writers in our corpus opened up dialogic space, they used more Entertain realisations than Attribute realisations. This may be due to their desire to focus on discussion of their own research (rather than to over-reference the research of others) combined with their awareness that readers would expect strong claims to be expressed with a degree of dialogic awareness.

Comparisons between sub-corpora

Having looked at the overall picture of evaluative language in the sub-genre investigated, we now turn to compare the two sub-corpora of L1 Chinese and L1 English writers to see whether there are any systematic differences between them. To do this we first normalised raw counts to frequencies per 1,000 words.

The figures below present comparison at different levels of delicacy; Figure 5 looks at Engagement overall, Figure 6 divides Engagement into Monogloss and Heterogloss, and Figure 7 compares all sub-categories of Heterogloss.

In this section we discuss similarities and differences between the normalised frequencies in the two sub-corpora, but it is important to note from the outset that none of the differences that we found proved to be statistically significant. The implications of this will be discussed in the final section of the article.

Figure 5 Normalised frequencies (per 1000 words) of Engagement features across two sub-corpora

Figure 6 Normalised frequencies (per 1000 words) of Monogloss and Heterogloss features across two sub-corpora

Figure 5 shows that overall, the L_1C sub-corpus contains slightly more instances of Engagement than the L_1E sub-corpus. Then as can be seen from Figure 6, heteroglossic options were more often used than the monoglossic option (Bare Assertion) across both sub-corpora, although the difference between the number per 1,000 words of Bare Assertion and Heterogloss in the L_1C corpus (n=9.5) is smaller than such difference in the L_1E corpus (n=13.67). This result suggests that the L1 English writers seemed to be slightly more conscious of engaging with alternative voices than did the L1 Chinese writers.

Figure 7 *Normalized Frequencies (per 1,000 words) of Sub-categories of Heterogloss across Two Sub-corpora*

Figure 7 above shows that, within the contractive subsystem, Disclaim was preferred to Proclaim by both the L1 Chinese and L1 English writers, which is in contrast with previous research which suggests that L1 Chinese writers are particularly likely to make strong, assertive statements (Chen, 2010; Hu & Cao, 2014; Hyland & Milton, 1997). It is however in line with research into other types of academic writing, for example Lancaster's (2011) finding that Disclaim rather than Proclaim was the more frequently used contractive option in the data (four argumentative texts in Economics) which he investigated.

As for expansive resources, Entertain was the most frequently employed option by both the L1 Chinese and L1 English writers. This is again in contrast with the suggestion from previous research that Chinese writers tend to intensified assertions in argumentative essays. However, it is in line with tendencies observed in some broader research, for example Fryer's (2013) finding that Entertain had the highest frequency of occurrence in medical research articles and showed the highest density in discussion sections of those articles, or Swain's (2010) identification of Entertain as the most widely used option in discussion essays written by undergraduate students.

The number of Distance categorisations per 1,000 words in the L_1C sub-corpus (n=0.44) is more than twice that in the L_1E sub-corpus (n=0.2). The finding that the L1 Chinese writers used more Distance, thus explicitly disalinging from sources, appears surprising in the light of previous research (e.g. Hood, 2004; Bitchener & Basturkmen, 2006) which argues that Chinese students are in general reluctant to critique. The L1 Chinese writers' preference for Pronounce indicates that, compared with their counterparts, they seemed to be more inclined to directly intervene in the text by presenting themselves as responsible for the new proposition being advanced. It is worth noting that Wu (2007), who compared higher and lower rated essays, also found a difference regarding Pronounce, with a preference for it being a characteristic of the lower-rated set.

A final interesting point of comparison is in the techniques used by each group to bring in external support for their assertions. Figure 7 shows that the L1 Chinese writers used more Pronounce, Justify-from-data, and Distance than did the L1 English writers, but the former group used less of the rest of heteroglossic options than did the latter group. The L1 Chinese writers' more frequent use of Justify-from-data suggests that they seemed to prefer using reference to the research participants' voice to support their assertions. An example from the corpus is:

12.   None use the checklist method discussed in section 6.2.1. This might be attributed to many teachers not being aware of the existence of published lists of criteria. ***All but T3 in the interviews reported that*** they had not taken any training course in textbook evaluation. (Text 1: 199)

In contrast, the L1 English writers' more frequent use of Endorse seems to indicate that this group of writers preferred referring to literature for the support of authorial assertions.

13.   In a discussion of the difficulties inherent in attempting complex change, ***Fullan (1991a) suggests that*** "the answer seems to be to break complex changes into components and implement them in a divisible and/or incremental manner" (p. 72). This seems to be in line with teacher C's suggestion in 5.4.1 that for TOC it is preferable to implement the teaching part before changing the assessment aspects. (Text 7: 277)

All the comparisons discussed above have shown differences in using some Engagement options by the two groups of writers – but, as was stated above, none of the differences proved to be statistically significant. We tested for statistical significance using the Mann-Whitney U test, in accordance with two principles suggested by Gries (2014): the type of study being conducted and the type of variables involved. The study compares two independent data sets; there is no relationship between the observations of the use of Appraisal options in each of the L_1C and L_1E sub-corpora or between the two sub-corpora. Therefore, our research is a type of study which is designed to describe the 'differences among texts and text varieties' (Biber & Jones, 2009: 1290). Each text is treated as 'an observation', and the 'rates of occurrence' of linguistic features are interval variables that can be subjected to inferential statistics (ibid). Given the small corpus size, a normal

distribution of data was not assumed, making a non-parametric test appropriate. (Field, 2009; Oakes, 1998). The Mann-Whitney U test satisfies all our requirements, and so a two-tailed test (p<0.05) was run on all data sets.

Test results for Engagement overall showed that the normalized frequency of Engagement in the L_1C sub-corpus (*Mdn* = 44.39) did not differ significantly from that in the L_1E sub-corpus (*Mdn* = 43.37), *U* = 16.00, *p* = 0.818, *r* = -0.09 (see Table 1 in Supplementary Data file). This result is at odds with many previous studies which observed differences in using interpersonal language by L1 Chinese and L1 English writers, albeit at lower education levels (e.g. Chen, 2010; Hinkel, 1997, 2003; Hyland & Milton, 1997). It suggests that, as language proficiency increases and experience in academic education accumulates, L1 Chinese writers, at least those in this study, have taken good command of making interpersonal meanings in thesis writing.

The statistical test also confirmed that there is no significant difference in the use of Bare Assertions and Heterogloss by the L1 Chinese writers and the L1 English writers, *U* = 10.00 and 17.00, *p* = 0.240 and 0.937, *r* = -0.37 and -0.05, respectively (see Table 2 in Supplementary Data file). As noted above, Swain's (2010) research showed that non-native undergraduates generally experienced more difficulty with effective deployment of Engagement resources. However, the L1 Chinese doctoral students in our study seemed to show similar rhetorical behaviour in using Engagement to their L1 English counterparts, particularly in terms of the relative frequency. Furthermore, the similar use of Heterogloss seems to indicate both groups of writers' awareness of engaging with the reader and construing dialogic divergences in their texts.

Conclusion

Our overall conclusion, that the difference in the patterns of use of Engagement resources by the two groups is not statistically significant, refutes the view that Chinese students are reluctant to critique and supports Pilcher, Cortazzi, and Jin's

20

(2011: 308) argument against the assumption that all Chinese students have some 'key traits of 'Chineseness'' such as 'preserving and maintaining public image'. In fact, the finding that the L1 Chinese writers in this research employed a large amount of Entertain counters previous views which have ascribed Chinese writers' tendency to overuse of strong assertions to the notion that Chinese rhetoric favours such a style (e.g. Chen, 2010). The evidence from our own small and context-specific dataset suggests that at the highest level of education, writers' first language may not exert as much impact on academic writing as it arguably does when writers are at a lower level. The L1 Chinese writers in our study had developed a good command of resources to evaluate others' research and to make their own propositions in a way that their academic discourse community finds acceptable. The L1 Chinese and the L1 English groups both wrote their theses in the same discipline and in the same academic department. It would seem that through this process, both groups became 'academically literate' (Lea & Street, 1998: 158) in terms of deploying interpersonal language.

Previous research has also suggested that local institutional culture may influence writers' deployment of interpersonal resources; for example Li & Wharton (2012) found that context, rather than discipline, was an influencing factor in writers' metadiscourse choices. Dahl (2004) argues that patterning of textual metadiscourse is influenced both by disciplinary culture and by L1 writing culture, but also shows that disciplinary culture can transcend national culture in research writing in disciplines such as Medicine with a stable and homogeneous knowledge base and structure for research reporting. Our study intentionally selected theses from a single discipline in order to avoid the possible complication of the influence of different disciplinary epistemologies. It therefore cannot be generalized to other disciplines, but it does indicate that the discipline under study may have been a homogenizing factor.

Given the limitations of our own study, a number of future research possibilities might be suggested. For example, future research could examine other aspects of Appraisal as well as Engagement, and could look at discussion sections of theses

21

from other disciplines in Social Science. It may also be interesting to compare Appraisal features in doctoral theses by L1 English and L1 Chinese writers in the same discipline as investigated here, but from different institutions, in order to see whether institutional provenance seems to be an influencing factor.

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*. Rather, our findings suggest that competence in this area of language use is developed via the PhD journey as undertaken within a specific institutional context.

PhD students are, after all, typically exposed to a large amount of generic modelling and associated critical discussion. They read past theses, write documents (such as interim research reports) that are in some senses thesis-like, and discuss drafts with their supervisors and other academic staff. Supervisors discuss language choices in drafts with all students, not only with those writing in their L2. We would suggest that this process, undergone by all twelve writers in our corpus, has led all of them to an appropriate genre-specific command of evaluative language.

References

Becher, T. (1990). The counter-culture of specialisation. *European Journal of Education, 25*(3), 333-346.

Biber, D., & Jones, J. (2009). Quantitative methods in corpus linguistics. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics* (Vol. 2): *An international handbook* (pp. 1286-1304). Berlin: Walter de Gruyter.

Bitchener, J., & Basturkmen, H. (2006). Perceptions of the difficulties of postgraduate L1 thesis students writing the discussion section. *Journal of English for Academic Purposes, 5,* 4-18. http://dx.doi.org/10.1016/j.jeap.2005.10.002

Brooke, M. (2014). Attribution and authorial (dis)endorsement in high- and low-rated undergraduate ESL students' English academic persuasive essays. *English Linguistics Research, 3*(1), 1-11. doi:10.5430/elr.v3n1p1

Bruce, I. (2014). Expressing criticality in the literature review in research article introductions in applied linguistics and psychology. E*nglish for Specific Purposes, 36,* 85-96. http://dx.doi.org/10.1016/j.esp.2014.06.004

Bunton, D. (1999). The use of higher level metatext in Ph.D theses. *English for Specific Purposes, 18,* S41-S56.

Carter, S. (2011). Doctorate as genre. *Higher Education Research and Development, 30*(6), 725-736.

Carter, S., & Blumenstein, M. (2011). Thesis structure: Student experience and attempts towards solution. In K. Krause, M. Buckridge, C. Grimmer & S. Purbrick-Illek (Eds.) *Research and development in higher education: Reshaping higher education*, *34* (pp. 95–107). Gold Coast: Higher Education Research and Development Society of Australasia, Inc.

Chen, H. (2010). Contrastive leaner corpus analysis of epistemic modality and interlanguage pragmatic competence in L2 writing. *Arizona Working Papers in SLA & Teaching, 17,* 27-51. Retrieved from http://www.coh.arizona.edu/AWP/Volumes.htm

Chatterjee, M. (2008). Textual engagement of a different kind? *Bridging Discourse: ASFLA 2007 Online Proceedings* (pp. 1-15). Australia: Australian Systemic Functional Linguistics Association.

Dahl, T. (2004). Textual metadiscourse in research articles: A marker of national culture or of academic discipline? *Journal of Pragmatics, 36,* 1807-1825.

Dudley-Evans, T. (1998). Introduction. In I. Fortanet, S. Posteguillo, J. Palmer, & J. Coll (Eds.), *Genre studies in English for academic purposes* (pp. 9-14). Castellón de la Plana: Publicacions de la Universitat Jaume I.

Field, A. (2009). Discovering statistics using SPSS (3rd ed.). London: SAGE Publications Ltd.

Flowerdew, L. (2004). The argument for using English specialized corpora to understand academic and professional language. In U. Connor & T. Upton (Eds.), *Discourse in the professions*: *Perspectives from corpus linguistics* (pp. 11-36). Amsterdam: John Benjamins Publishing Company.

Gabrielatos, C., & McEnery, T. (2005). Epistemic modality in MA dissertations. In Fuertes Olivera, Pedro A. (Ed.), *Lengua y sociedad: Investigaciones recientes en lingüística y filología no. 61*(pp. 311-331). Valladolid: Universidad de Valladolid.

Gholami, J., Nejad, S., & Pour, J. (2014). Metadiscourse markers misuses: A study of EFL learners' argumentative essays. *Procedia - Social and Behavioral Sciences, 98,* 580-589. doi:10.1016/j.sbspro.2014.03.454

Gries, S. (2014). Basic significance testing. In R. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (pp. 1-19). Cambridge: Cambridge University Press. Retrieved from http://www.linguistics.ucsb.edu/faculty/stgries/research/ToApp_STG_BasicSignTesting_ResMethLing.pdf

Lancaster, Z. (2011). Interpersonal stance in L1 and L2 students' argumentative writing in Economics: Implications for faculty development in WAC/WID programs. *Across the Disciplines, 8*(4). Retrieved from http://wac.colostate.edu/atd/ell/lancaster.cfm

Liu, X., & Thompson, P. (2009). Attitude in students' argumentative writing: A contrastive perspective. *University of Reading Language Studies Working Papers, 1,* 3-15.

Hinkel, E. (1997). Indirectness in L1 and L2 academic writing. *Journal of Pragmatics, 27,* 361-386.

Hinkel, E. (2003). Adverbial markers and tone in L1 and L2 students' writing. *Journal of Pragmatics, 35,* 1049-1068.

Hood, S. (2004). *Appraising research: Taking a stance in academic writing* (Unpublished doctoral thesis). University of Technology, Sydney.

Hood, S. (2005). Managing attitude in undergraduate writing: A focus on the introductions to research reports. In L. Raveli & R. Ellis (Eds.), *Analysing*

academic writing: Contextualized frameworks (pp. 24-44). New York, NY: Continuum.

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

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hyland, K. (2010). *Metadiscourse: Mapping interactions in academic writing. Nordic Journal of English Studies, 9*(2), 125-143. Retrieved from http://hdl.handle.net/10722/130125

Hyland, K., & Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing, 6*(2), 183-205.

Koutsantoni, D. (2006). Rhetorical strategies in engineering research articles and research theses: Advanced academic literacy and relations of power. *Journal of English for Academic Purposes, 5,* 19-36. doi:10.1016/j.jeap.2005.11.002

Lea, M., & Street, B. (1999). Student writing in higher education: An academic literacies approach. *Studies in Higher Education, 23*(2), 157-172. Essex: Addison Wesley Longman Limited. http://dx.doi.org/10.1080/03075079812331380364

Li, T., & Wharton, S. (2012). Metadiscourse repertoire of L1 Mandarin undergraduates writing in English: A cross-contextual, cross-disciplinary study. *Journal Of English For Academic Purposes*, *11*(4), 345-356. http://dx.doi.org/10.1016/j.jeap.2012.07.004

Martin, J. R., & White, P.R.R. (2005). *The language of evaluation: Appraisal in English*. New York, NY: Palgrave Macmillan.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. New York, NY: Routledge.

Paltridge, B., & Starfield, S. (2007). *Thesis and dissertation writing in a second language: A handbook for supervisors*. New York, NY: Routledge.

Parkinson, J. (2011). The Discussion section as argument: The language used to prove knowledge claims. *English for Specific Purposes, 30,* 164-175. doi:10.1016/j.esp.2011.03.001

Petrić, B. (2007). Rhetorical functions of citations in high- and low-rated master's theses. *Journal of English for Academic Purposes, 6,* 238-253. doi:10.1016/j.jeap.2007.09.002

Pilcher, N., Cortazzi, M., & Jin, L. (2011). Different waves crashing into different coastlines? Mainland Chinese learners doing postgraduate dissertations in the UK. In L. Jin & M. Cortazzi (Eds.), Researching cultures of learning: International perspectives on language learning and education (pp. 292-313). Hampshire: Palgrave MacMillan.

Swain, E. (2010). Getting engaged: Dialogistic positioning in novice academic discussion writing. In E. Swain (Ed.), *Thresholds and potentialities of systemic functional linguistics: Multilingual, multimodal and other specialised discourses* (pp. 291-317). Trieste: EUT Edizioni Università di Trieste. Retrieved from http://hdl.handle.net/10077/3647

Swales, J. (2004). *Research genres: Explorations and applications*. Cambridge: Cambridge University Press.

White, P. (2001). Engagement. In *Appraisal*: *An overview*. Retrieved from http://grammatics.com/appraisal/AppraisalGuide/Framed/Frame.htm

White, P. (2012). Exploring the axiological workings of 'reporter voice' news stories—Attribution and attitudinal position. *Discourse, Context & Media, 1,* 52-67.

Wu, S. M. (2007). The use of engagement resources in high- and low-rated undergraduate geography essays. *Journal of English for Academic Purposes, 6,* 254-271. doi:10.1016/j.jeap.2007.09.006

Appendix: Texts in the corpus

|	Word length of text	Number of Engagement features	Engagement features per 1,000 words	Writer's first language
Text 1	6059	292	48.2	Chinese
Text 2	13041	658	50.5	Chinese
Text 3	8127	371	45.7	Chinese
Text 4	9376	393	42.0	Chinese
Text 5	12779	552	43.2	Chinese
Text 6	4849	183	37.7	Chinese
Text 7	7875	381	48.4	English
Text 8	12733	529	41.5	English
Text 9	10232	555	54.2	English
Text 10	13620	610	44.8	English
Text 11	10775	378	35.1	English
Text 12	9505	398	41.9	English

Tables and Figures
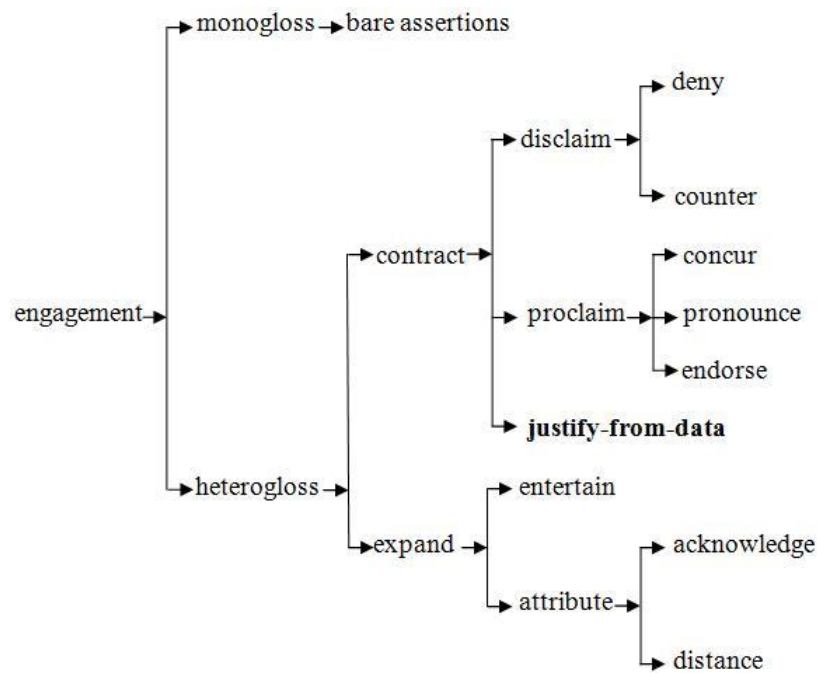
Figure 1 Analytical framework used in this study, adapted from Martin & White (2005)

"The freedom in learning that learners can exercise has always been a major concern of many autonomy advocates (e.g. Benson, 2001; 2006a). For example, **Benson (2001) emphasised** important aspects of such freedom as being the free choice of the learning contents and of the learning process."

[Analyst comment] *An example of Engagement: Proclaim: Endorse?*

[Co-author comment] *So hard to tell – as you read ahead, do you see that the author does align strongly with Benson?*
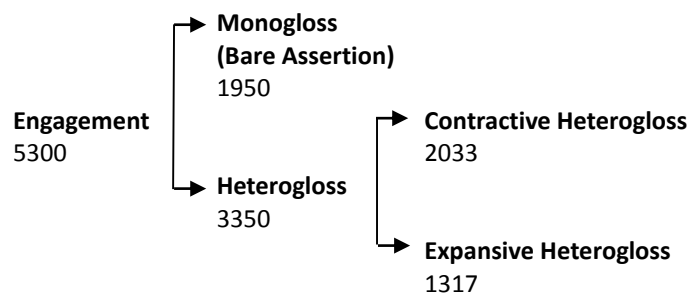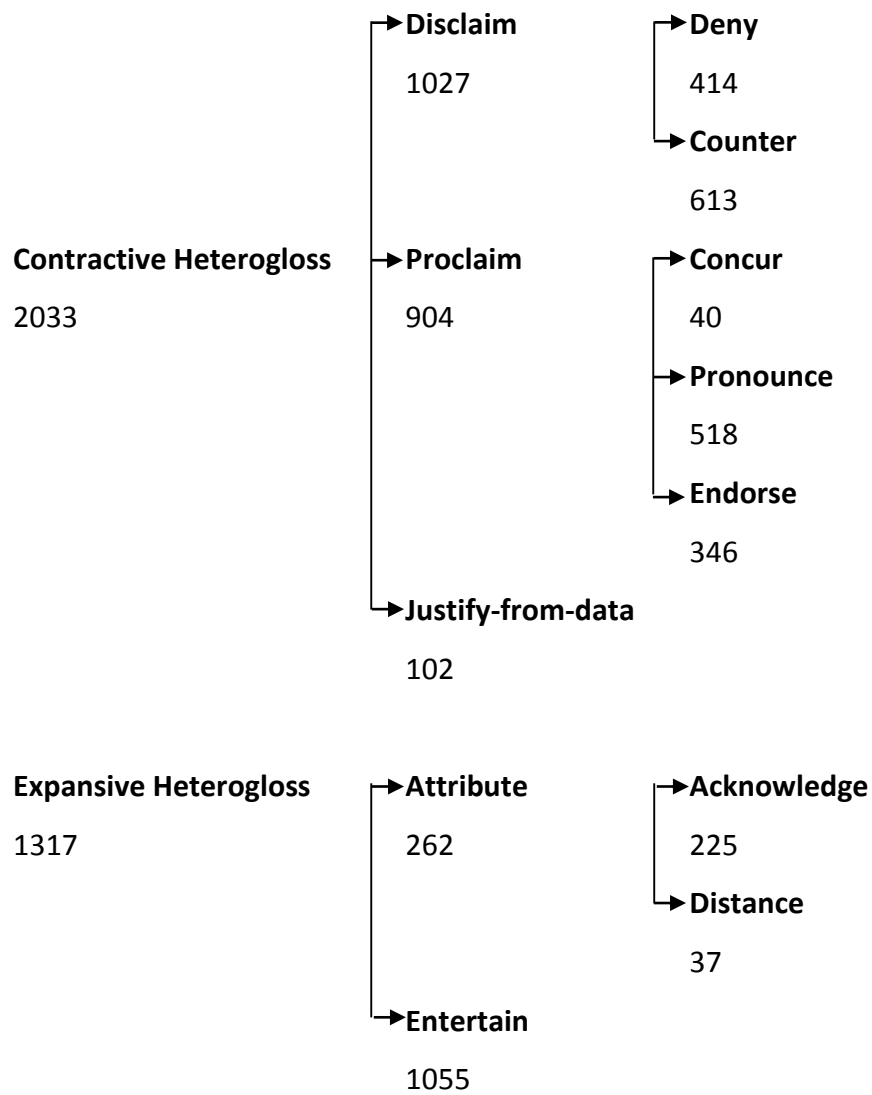
Figure 2 Extract from analyst notes

Engagement
5300

Monogloss
(Bare Assertion)
1950

Heterogloss
3350

Contractive Heterogloss
2033

Expansive Heterogloss
1317

Figure 3 Engagement in the whole corpus

Disclaim

1027

Deny

414

Counter

613

Contractive Heterogloss

2033

Proclaim

904

Concur

40

Pronounce

518

Endorse

346

Justify-from-data

102

Expansive Heterogloss

1317

Attribute

262

Acknowledge

225

Distance
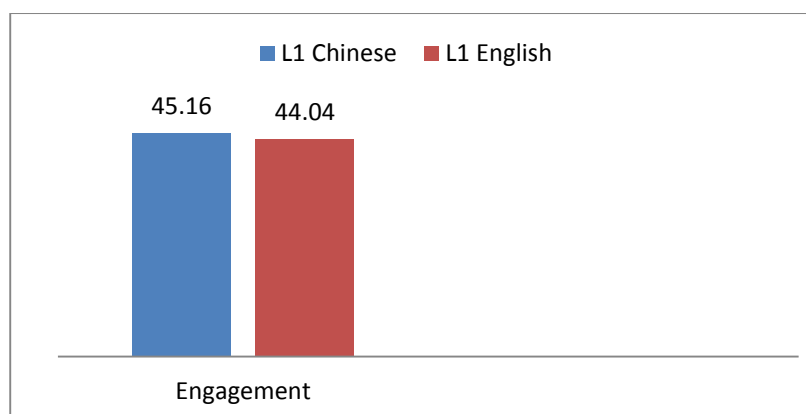
37

Entertain

1055

Figure 4 Heterogloss Features in the Whole Corpus

Figure 5 Normalised frequencies (per 1000 words) of Engagement features across two sub-corpora
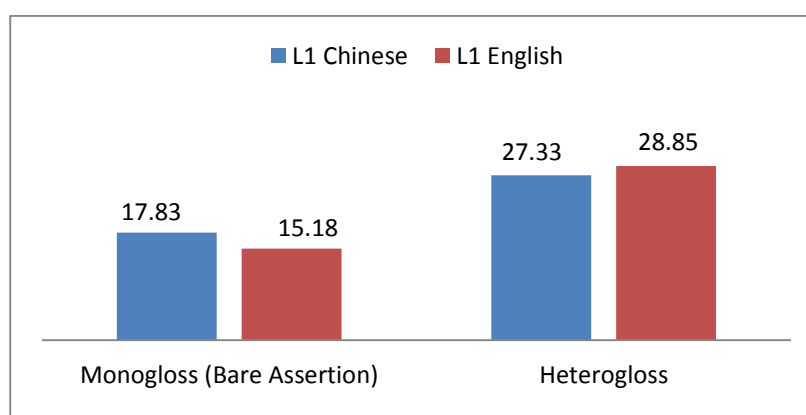

Figure 6 Normalised frequencies (per 1000 words) of Monogloss and Heterogloss features across two sub-corpora
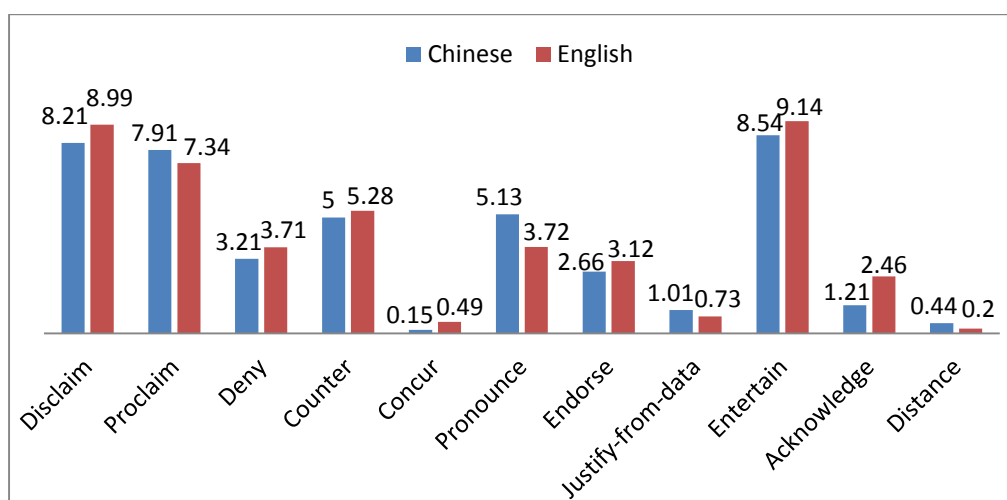

Figure 7 *Normalized Frequencies (per 1,000 words) of Sub-categories of Heterogloss across Two Sub-corpora*