Responses to Reviewers

We would like to extend our gratitude to the reviewers for their detailed and instructive comments, and we hope that we have sufficiently addressed the concerns raised. We believe the revised manuscript is much clearer and stronger due the comments provided. Listed below are changes made in response to reviewers' comments.

REVIEWER #1

1) Overall a very interesting, well designed, and well written study. However, one of the major concerns is that the authors are comparing L1-English senior-level students and L1-Chinese university students. Thus, they cannot say they are comparing the writing of peers. Much development happens within the college years even for L1 English writers (see, e.g., Staples, Egbert, Gray, & Biber, 2016). In addition, there would be differences expected in bundle use across these two contexts even for the same writers as the registers/genres written in the senior year are very different from FYW genres, and we also know that there are differences in bundle use across disciplines (e.g., Hyland, 2008). The authors need to acknowledge this gap in the data from the beginning of the study and throughout.

Response: Thank your raising this issue. We added an "extensive" discussion concerning this issue in the Methods and Conclusion sections. Please see the Methods section (p. 8) and Conclusion section (p. 29-31).

2) In addition, there are some points of clarity needed in the methods. Most notably, the application of the functional error coding should be described in more detail and justified.

Response: We revised the section on misuses in the Methods section to make it clearer as to how the analysis was conducted. Please see the Methods section (p. 15).

p. 9: is the second draft the final draft? Okay—p. 10 clearly indicates that this is the middle draft and why this draft was chosen. Might just mention something here at the first mention of the second draft.

Response: To avoid confusion, we just stated that the essays were the final assessed, sourced-based argumentative essay assignment in the first mention of the ESL essays, since we indicate that the essays were the middle draft and why they were selected in the subsequent mention. Please see the Methods section (p. 10).

p. 11: could the authors clarify and add information in the draft on whether the bundles were identified across both sets of texts or whether different bundles were identified in the two corpora? In other words, was the cut off 40 per million words using the entire data set (both corpora)? It seems that this was NOT the case, but this could be clarified. It would also be helpful to justify why this method was used (rather than pooling the data), since both approaches have been used.

Response: Actually, the cut-off frequency of 40 per million words was set for both corpora. We clarified this by adding "for both corpora in order to compare the bundles in these two corpora of different sizes." Please see the Methods section (p. 12).

p. 14: Can the authors clarify how they identified misuse and how this relates to the overall bundle counts? For example, if a writer used on other hand instead of on the other hand, the bundle would not show up in the frequency analysis, right? So how would this misuse be found? Did the researchers search content words within the bundles to find these misuses? Or some other method?

Response: We revised the section on misuses in the Methods section to make it clearer as to how the analysis was conducted. Please see the Methods section (p. 15).

p. 15: results for number of bundle tokens that meet the threshold- I think that this may be influenced by the topics/genres/disciplines represented. The L2 corpus is a much more homogenous corpus than the L1 corpus. One would expect more similarities across the same genre/context of use than across the more heterogeneous MICUSP corpus. Something should be added here and/or in the discussion to address this.

Response: Thank you for bringing this to our attention. We have added the following (pp. 17-18): "It is also likely that the quantitative differences may be due to the composition of the two corpora. While CXLTE is a much more homogenous corpus consisting of argumentative essays on general topics in the content of a non-disciplinary writing course, the MICUSP data used in this study is a heterogeneous corpus comprised of discipline-specific argumentative texts. As Hyland (2008a, 2008b) has shown, bundles are intimately connected to genres and disciplines." We also provided commentary on this issue in the Conclusion section (p. 29).

p. 16: It would be helpful to move the discussion of log-likelihood to the methods section, so that readers are prepared for this statistical test. When this is done, please also clarify whether the statistical test was conducted using the token frequency or type frequency.

Response: We moved the discussion of log-likelihood to the Methods section and clarified that we performed the tests on the token frequencies. Please see the Methods section (p. 16).

p. 19: Table 7—it is interesting that the use of referential bundles was similar for the L1 and L2 writers. I'm wondering if it would be possible to compare the totals by the broader category (e.g., stance, discourse organizing, referential) in this table so that readers can get a sense of the overall functional use. I see that the authors have included this in table 8, but I would suggest combining the information.

Response: Table 7 (p. 21) now includes the total frequencies of the broader functional categories. We also did this for Table 5 (p. 19)—structural categories. However, we kept Table 6 (p. 20) and Table 8 (p. 22) that report the proportional distributions of the broader structural and functional categories, as we find them still useful in showing the different patterns that emerge when we look at the percentages.

In addition, I'm assuming that the bundles in parentheses in Table 7 are examples, but it would be helpful to add e.g. to make this clear.

Response: We added *e.g.* before each example in Table 7 as well as in Table 5. Please see p. 19 (Table 5) and p. 21 (Table 7).

p. 22: I'm a bit confused as to why the authors switched from log-likelihood to chi square here. Log-likelihood has become more preferred for comparison between groups across frequency of occurrence (when not using means based tests). As with the log-likelihood test, the use of chi-square should be described in the methods and justification for using each method should be provided.

Response: We decided to delete the chi-square tests, as they do not really seem to show anything interesting.

p. 25: "a large amount of"◊ I'm confused as to how this is a functional error. The writer is still using the bundle to provide information about quantity, so the error seems to be a grammatical one. As the authors point out, a good substitute would be a large number of. This would correct the agreement with number, a grammatical error not a functional error.

Response: The reviewer is correct. This was a mistake on our part. It should have been classified as a grammatical mistake—agreement. We have revised that part. Please see p. 27.

p. 26: "according to the article"- again, I see this as a function of grammar/syntax. This is forming a kind of topic-comment structure that is common in developing L1-Chinese writers' texts (and this particular error is a familiar one). I don't understand the classification of this as a functional error. For both this case and the one above, it would be helpful to describe more fully in the methods why these are classified as functional errors. They both seem different from the misuse of "on the other hand" found in excerpt 1.

Response: Thank you for bringing this to our attention. After careful consideration, we have come to agree with the reviewer. Initially, we classified it as a functional mistake because, in Huang (2015), she presents a similar redundancy and seems to have classified those instances as functional mistakes, at least in writing. However, after reading the reviewer's comment and other sources on L1-Chinese student writing (e.g., Li, 2017), we decided to follow the reviewer's suggestion and classify this as a grammatical mistake. Similar to the reviewer, Li reports that this type of topic-comment structure is common in developing L1-Chinese writers' academic text. We revised that part (p. 27) and updated Table 12 (pp. 25-26) to reflect the changes as well.

p. 27: As mentioned above, the discussion needs to account for the fact that differences are also likely a result of genre/disciplinary differences in the texts analyzed.

Response: We added the following in the conclusion (p. 29): "Furthermore, as Hyland (2008a, 2008b) has shown, the context of writing (e.g., genre, discipline, author, audience) has a crucial effect on bundle usage. This study compared L1-Chinese argumentative essays written within an

FYC course for international students with essays written by L1-English senior undergraduates for disciplinary courses. It is likely that disciplinary reader expectations played an important role in shaping the writing, including the use of bundles, of the L1 texts selected, while such disciplinary expectations may not have influenced the ESL essays."

p. 28: Here the authors provide some support for why it might be helpful to compare senior-level L1 writing texts. However, it still seems unfair to expect L2 students to write at this level when their L1 peers are almost certainly not at this developmental stage.

Response: We added an "extensive" discussion of this concern in the Conclusion section. Please see the Conclusion section (pp. 29-31)

REVIEWER #2

1) I feel there is an important theoretical or epistemological flaw here -- not exclusive to this manuscript and indeed common in this type of research -- that I would like the authors to address: why is it worthwhile at all to compare L1 and L2 academic writing? It can no longer be assumed, nor should it, that "native" English writers and their written production are a logical target for L2 English writers. I know that it lies beyond the scope of the authors' study to address issues of "ELF" and similar, but it should also not simply go unaddressed at all. After all, the authors are comparing "highly rated" papers produced by native writers of English to other "highly rated" papers written by non-natives. Put another way, the result seems to be the same: both groups being compared met with success, as measured by getting good grades on their work. So why change?

Response: Thank your raising this issue. We added an "extensive" discussion concerning this issue in the Methods and Conclusion sections. Please see the Methods section (p. 8) and Conclusion section (p. 29-31).

2) Yes, there is an argument to be made that the CXLTE corpus shows on overuse of adjectives in stance bundles (thus potentially diverging from what is generally considered good academic register), but why is the word "overreliance" used by the authors when describing the relative abundance per se of lexical bundles used by Chinese L1 speakers? More bundles = bad? The morpheme "over" to me carries a negative connotation - why the negativity?

Response: We removed expressions such as "overreliance," "overuse," "underuse," etc.; we instead replaced those expressions that carry potentially negative connotations with descriptive terms such as "more/less frequently," "more/less frequent," "greater/fewer," etc. throughout the manuscript.

3) This epistemological fallacy is further exemplified by the authors pointing out that there were many bundles attested in the CXLTE corpus that were absent in the MICUSP counterpart. That fact alone does not mean much, in my view, unless the authors are suggesting that if native speakers of English do not use those bundles then they are not correct, or in some way inferior. If that is what they authors believe, fine, but please make that case. If that is not what the authors believe, then why include such information?

Response: It was not our intention to suggest that they are not correct because they do not appear in MICUSP. We pointed these out to show that there are bundles that appear only in CXLTE. To avoid such interpretation, we decided to omit such statements throughout the manuscript.

Similarly, I wonder why the authors chose to analyze bundles for "misuse" only in the Chinese-speaker corpus. Is it impossible that there were also misuses present in the MICUSP data?

Response: We added the following in the Methods section to address this concern (p. 14): "While we acknowledge that it is possible that bundle misuses may also exist in advanced L1 student writing, no such misused instances were found in the bundles identified in the MICUSP data used. We, therefore, concentrate exclusively on bundle misuses in the L1-Chinese writers' texts in order to identify categories of multiword sequences that these groups of learners may need better control over in academic writing."

4) - On p.3 the authors mention that Biber suggests that lexical bundles may be retrieved from memory as "chunks". Psycholinguistic claims of this nature regarding lexical bundles has never been a main contention of Biber's, and I would avoid making this sort of assertion in this paper.

Response: We deleted that sentence.

5) - I believe it is relevant to mention where "XXX University" (p.9) is located (which country), as 1st year comp in the USA (MICUSP) may not be comparable to other first year composition courses elsewhere.

Response: We clarified that the location of FYC course is the US in the Introduction section, and, in the Methods section (p. 9), we indicated that the XXX University is in the US Midwest.

6) - I'm not sure I'm comfortable with the conflation of the Results and Discussion sections in this paper. I would rather see the Results speak for themselves first, allowing the reader to make her or his own inferences, and then, separately, hear the authors' take on the results. To do otherwise, as is currently the case in this submission, puts the results in the background without giving the reader a chance to first consider and digest the data that arose from the study, forcing the reader to read the results through the filter of the point the authors want to make. In some articles this can work fine, but in this case, in which there are important native/non-native issues at play, I feel it is best to separate the sections.

Response: Thank you for the suggestion, but we feel that combining the results and discussion makes it easier to read this manuscript. We offer more commentary of the general findings and other matters, such as pedagogical implications, in the Conclusion section.