

Informing Academic Writing Pedagogy through the Study of Phrase-frames

Kseniya Walcott

Colorado State University/English, Fort Collins, United States

Abstract—Second Language Development (SLD) research has investigated aspects of complexity, accuracy, and fluency (CAF) to understand how English language learners (ELLs) develop linguistic faculties related to language production, especially in academic writing (Biber & Gray, 2010; Biber, Gray & Ponpon, 2011; Kyle & Crossley, 2018; Lu & Ai, 2015). Contributing to the current body of SLD research, this study addresses complexity in expert academic argumentative essays by investigating language at the phrasal level through phrase-frames (PFs). Phrase-frames are fixed or semi-fixed syntactic structures with a variable slot that might be considered another measure of syntactic complexity. This study aimed to identify if native and non-native expert writers in the Michigan Corpus of Upper-Level Student Papers (MICUSP) displayed significant differences in their use of these frames concerning their predictability, variability, and function. Results of the study indicated expert academic writers in this corpus use highly variable, highly unpredictable structures which confirm the complexity, elaboration, and explicitness unique to the academic register. Understanding of these register features through the identified PFs was used to develop sample pedagogical materials to promote more complex writing in academic genres.

Index Terms—second language development, phrase-frames, syntactic complexity, MICUSP, EAP

I. INTRODUCTION

A significant amount of psycholinguistic research asserts that language is composed of fixed and semi-fixed structures which, if learned by students, can help lighten their cognitive load during language production (Garner, 2016). Recent studies have focused on identifying fixed and semi-fixed structures in learners' language production within the dimensions of complexity, accuracy, and fluency (CAF) to better understand how these elements interact with each other and their use in measuring learner progress (Vercellotti, 2017; Yang & Sun 2015). Of particular interest has been the examination of complexity in a learner's language through their syntax in the form of phraseological units. Thus, phrase-frames (PFs) come into the picture as another potentially fruitful unit of study for phraseological development in learner language that, perhaps, could be considered among the many indicators of complexity in language.

While many studies utilize the term Second Language Acquisition (SLA), the current study employs the term Second Language Development (SLD) in its place as it more closely aligns with the stance of language as a complex developmental *process* rather than product to be acquired by the learner (Bot, Lowie, Thorne, and Verspoor, 2013). In SLD research, careful examination of successful employment of syntactic complexity through the unit of phrase-frames in various genres may prove to be a potentially worthwhile exploration in determining how expert interlocutors engage with these structures. From this knowledge, pedagogical materials could be developed for English language learners (ELLs) especially in more difficult genres of communication (such as the argumentative essay) for a variety of contexts.

The current study incorporated research of PFs in highly rated argumentative essays composed by both native and non-native writers in the Michigan Corpus of Upper-Level Student Papers (MICUSP) as it provides a strong compilation of highly rated expert texts. Written academic argumentative essays were the genre of choice, because, compared to other written genres, they seem to be one of the most difficult to learn due to the many elements a learner must consider in the composition process, such as their language-specific nature and the variety of argument structures that could be considered successful or unsuccessful (Kaewpet, 2018). This study's primary focus was to determine if there were significant differences in syntactic complexity among expert writers' use of PFs in academic writing, examining variability, predictability, and functional features. Once these features were examined, sample pedagogical materials were developed with the intent of making these structures more salient to students in an English for Academic Purposes (EAP) setting, furthering the development of their academic writing.

Following this section, there is a discussion of how previous research concerning syntactic complexity has informed SLD writing pedagogy with a stress on the phraseological dimension. After, phrase-frames as a potential phraseological unit of complexity and previous research regarding these units in the EAP and English for Specific Purposes (ESP) contexts are addressed. Finally, recent corpus research about PFs and L2 writing conducted with MICUSP conclude the literature review, leading into the details of the current study.

II. LITERATURE REVIEW

A. Syntactic Complexity as a Crucial Dimension for Studying SLD Writing

Traditionally, when observing second language development in learners' writing, the three dimensions demanding diligent study have been complexity, accuracy, and fluency (CAF). As asserted by Larsen-Freeman (2009) and many others, these key factors in the SLD process are inseparable from each other working as a symbiotic system so that when a learner works on an aspect of complexity, accuracy and fluency are also dually affected. This has been demonstrated by Yang and Sun (2015) who examined the CAF in five undergraduate multilingual learners' writing throughout an academic year. From this study, they found CAF elements were mutually dependent on each other in the process of SLD and that L1, L2, and L3 writings consisted of emergent and constant changes, supporting the idea that language development is a complex, dynamic system impacting the learners' development in the target language over time. By examining these factors, a more holistic understanding of SLD can be attained, but measuring them remains a critical issue in the field, particularly concerning complexity in language development. As is the case with Yang and Sun (2015) and other research concerning the construct of complexity, it manifests and has been measured in many forms in learners' writing (syntactic, grammatical, lexical, morphological, etc.) including through the use of specific indices. Of the diverse forms of complexity, syntactic complexity seems to be most extensively explored in the field of SLD and can be defined as "...the range of syntactic structures produced [by learners] and the degree of sophistication of such structures" (Lu, 2011, p. 36). It has long stood as a measure of L2 learner development due to its prominence as a key feature in L2 writing pedagogy and research.

Syntactic complexity was considered, especially in earlier research, to be observable primarily based on clausal measures of complexity such as unit-length, subordination, and coordination; however, with the work of Biber and his colleagues, it was then suggested that phrase-level measures may more effectively identify trends in the development of academic writing and L2 writing in general (Biber et al., 2011). The idea that phrasal measures are the stronger indicator of writer development continues to be supported by recent research. For instance, in Kyle and Crossley's study (2018) they employed fine-grained indices of clausal and phrasal complexity to predict holistic scores of writing quality for a corpus of 480 independent essays from the TOEFL. The results of their study showed that the strongest indicators of writing quality came from phrasal rather than traditional or clausal fine-grained indices. It was also acknowledged, however, that a fine-grained clausal complexity index could be used as a supplemental tool alongside that of a phrasal one (Kyle & Crossley, 2018).

Learner proficiency, L1 background, gender, and the environments for L2 exposure are all factors (among many others) that have been carefully examined in conjunction with syntactic complexity indices to understand how the interaction of these elements could provide a glimpse into learners' performance and development in L2 writing situations (Kyle & Crossley, 2018; Lu & Ai, 2015; Martinez, 2017; Martinez, 2018; Yoon, 2017). In Martinez's (2018) study, she examined differences in syntactic complexity between lower-intermediate and intermediate secondary education writers with an additional focus on gender, finding that syntactic complexity increased in all aspects from the lower intermediate to the intermediate levels and that female participants generally surpassed their male counterparts concerning general quality in their compositions. Yoon's (2017) study, which analyzed a corpus of 1198 argumentative essays written by college-level Chinese EFL learners concurs with Kyle and Crossley's (2018): phrase-level syntactic complexity provides a more effective measure for learner proficiency in writing, since advanced writers more often express complex ideas through increased phrasal density, whereas there was little difference in clause-level complexity across proficiency levels. Lu and Ai's study (2015), which analyzed 1600 argumentative essays from students of different L1 backgrounds, again echoes the most current studies. Most advanced level writers employed phrase-level complexification in their works and the results also suggested that if they came from a certain L1 background, they may be more likely to produce more complex writing in English (though further research would be needed to confirm which L1 backgrounds might prove to be more advantageous) (Lu & Ai, 2015). Finally, Martinez's work (2017) explored syntactic complexity in the writing of secondary education students from bilingual and non-bilingual contexts finding that scores on all syntactic complexity measures were higher among the bilingual group who had more exposure to the target language. Therefore, based on the most recent studies concerning syntactic complexity, the phrasal aspect of syntactic complexity should also be considered in research on learner proficiency and environmental factors both internal and external which may impact a learner's writing.

B. Phrase-frames: Another Potential Measure of Syntactic Complexity

Due to its prominence in studies targeting syntactic complexity, phraseological complexity continues to be a worthwhile dimension to explore in the examination of ELL writing. It makes sense that phrasal complexity measures continue to be a focal point of much SLD research since many psycholinguistic studies confirm that in language development, understanding key fixed or semi-fixed sequences can lighten the cognitive load for learners so they can focus on other tasks (Garner, 2016). Phrase-frames (PFs) are fixed or semi-fixed multi-word sequences that "...allow for an element, often a word, to be 'plugged in' a standardized functional piece of language" (Garner, 2016, p. 32; Cunningham, 2017, p. 72). Some examples of this would include units such as: *it is* [clear, interesting, possible] *that*; on *the* [other, one] *hand*; and *as a* [result, part, matter] *of* where, as demonstrated by these examples from the corpus, a variety of different words (i.e., variants) could occupy the blank space within the frame.¹ This measure, along with other

¹ All examples of phrase-frames, unless stated otherwise, come directly from the corpus investigated in this study.

phraseological units (i.e., n-grams), have been used to establish the phraseological profiles of target texts regarding distributional, variational, and functional trends (Römer, 2010).

But successfully applying these structures can be a difficult endeavor, especially in academic writing. As Gray and Biber (2010) establish in their study, academic writing has a reputation for being elaborated, complex, and explicit so that academic language in writing is complex and elaborated regarding embedded noun phrases and explicit in highlighting characteristics of referents. Complexity and elaboration remain a prevalent feature in academic writing through past and current studies in phraseology, but particularly those focused on phrase-frames. Writers may need to address all these aspects when composing in the academic register. For instance, they may need to use different functional frames for explicitness through a high usage pattern of *specification of attributes* which work to identify referents; *metadiscourse markers* which would overtly express logical relations between ideas in the text; and *evaluation* and *epistemic stance* markers displaying the levels of accepted personal elaboration that are allowed in this genre of writing. Therefore, further phraseological research with a focus on PFs and their role in EAP pedagogy is needed.

There have been several recent exploratory studies investigating the significance of this structure and what it could indicate regarding discourse development in native and non-native pieces of writing both in the contexts of English for Specific Purposes (ESP) and English for Academic Purposes (EAP). For instance, concerning ESP research, Grabowski's (2015) study sought to identify the use, structural and functional features of PFs in four English pharmaceutical text types: patient information leaflets, summaries of product characteristics, clinical trial protocols, and chapters/sections from academic textbooks on pharmacology. Based on his analysis of 4-word units with an open slot in the medial position, he found that even within the same register, the use and discourse functions varied across different genres, justifying the treatment of the discourse functions of phrase-frames, as separate from their variants. Cunningham (2017) in her research on PFs analyzed a corpus of recent mathematics research articles finding that this genre had its own distinctive set of PFs which specifically characterize that kind of writing. She acknowledged after conducting this study that "...phrase-frames can provide students with a clearer understanding of how language works within a specific discipline, and such analysis can give instructors a start to understanding new disciplines and genres" (Cunningham, 2017, p. 81). Her study highlights the pedagogical potential of PFs as a measure to help learners and instructors alike distinguish the specific conventions of discipline-specific writing.

While Grabowski's focus was on natively composed texts, Garner's (2016) EAP research focused on analyzing learners' use of phrase-frames in their own writing. In his study, he observed phrase-frames produced by EFL L1 German learners across five different proficiency levels in the *EF-Cambridge Open Language Database* (EFCAMDAT) finding that writers at higher proficiency levels produced more syntactically varied, unpredictable and complex structures in their texts. In the same vein as Cunningham (2017), Lu, Yoon, and Kisselev (2018) took a pedagogically motivated PF approach in their study and compiled a list of "pedagogically useful" PFs. Their list consisted of research article introductions in six social science disciplines consisting of 370 five-word phrase-frames and 84 six-word phrase-frames. In order to determine pedagogically useful PFs, they had a panel of academic writing instructors and student writers evaluate a random sample of 100 PFs, and based on these evaluations, concluded that their identified frames were considered useful for instruction by either the instructors or the student writers, or both. Investigating PFs' role in the development in learner writing, Juknevičienė and Grabowski (2018) analyzed writing from the Polish and Lithuanian sub-corpora from the International Corpus of Learner English (ICLE) and writing from British and American students in the Louvain Corpus of Native English Essays (LOCNESS). The authors determined there were differences in PF use between the Polish and Lithuanian writings due to L1 transferal, but also noticeable similarities between the Lithuanian and Polish students who used some less frequent PFs or PFs that were completely absent from the native texts presented in the LOCNESS corpus. Finally, Nekrasova-Beker's (2019) study investigated PF structural and functional characteristics in pedagogical materials for undergraduate lower-division engineering courses. She found that PFs shared characteristics across engineering disciplines regarding position of the variant slot, and structural and functional features. However, the variants that could occupy the frame presented a high range of variability even within that single discipline, aligning with Cunningham's (2017) study which emphasizes that even within a respective discipline's discourse, there can be great variation in PF usage. All of these studies confirm once again that there are complex factors at play in the language development process for L2 writers which can be observed through PFs: L1 influence, common learner trends in PF usage, PF exposure through pedagogical materials, and distinguishing usage of PFs in native writing.

Thus, based on this expansive body of SLD research, it appears evident that language is a non-linear complex process that demands a great deal of cognitive processing (Garner, 2016). However, quicker processing of specific linguistic structures can be enabled by increasing learners' awareness of PFs in the classroom. The most recent studies in PFs have established that there are, indeed, differences in usage and function of PFs across genres and disciplines (Grabowski, 2015; Cunningham, 2017; Lu, Yoon, & Kisselev, 2018; Nekrasova-Beker, 2019), across learner proficiency levels (Garner, 2016), and between learners with a different L1 and native speakers (Juknevičienė & Grabowski, 2018). This study hopes to further contribute to the body of literature regarding how PFs can be identified from corpora of expert texts and successfully implemented into L2 writers' syntactic repertoire. The current study builds on the previous research on PFs while offering further insight into the variability, predictability, and functional

characteristics within the expert corpus of Michigan Corpus of Upper-Level Student Papers (MICUSP) argumentative essays. Based on the results of this study, the goal is then to develop and directly implement PF-focused pedagogical strategies in an EAP instructional setting to help learners enhance their academic writing.

C. *Brief Overview of MICUSP Corpus*

As stated above, this current study aims to unify research with practical application of the prominent structure of PFs in learner writing, drawing upon the methodology established in previous research. Argumentative essays were extracted from MICUSP, a compilation of A-graded papers from students at the University of Michigan from four different student levels (senior undergraduate-3rd year graduate student) and 16 different disciplines which include Biology, Civil & Environmental Engineering, Economics, Education, English, History & Classical Studies, Industrial & Operations Engineering, Linguistics, Mechanical Engineering, Natural Resources & Environment, Nursing, Philosophy, Physics, Political Science, Psychology, & Sociology (Römer & O'Donnell, 2011). This collection of texts spans seven academic written genre types: the argumentative essay, creative writing, critique/evaluation, proposal, report, research paper, and response paper. Of these seven genres, the argumentative essay is the second highest in distribution in the corpus at 22.4% only coming after the report at 43.9% (Römer & O'Donnell, 2011). Writers of these papers included native speakers of English who composed 82% of the corpus (681 papers) and non-native speakers of the language who composed 17.9% (148 papers) of the corpus. Since its compilation in 2009, MICUSP has served as a collection of exemplary academic written texts used to inform pedagogical materials development and research.

D. *Studies of Writing Using MICUSP and Other Corpora*

Since its creation, MICUSP has served as an informative corpus for L2 research concerning genre-specific characteristics of academic writing which have been used to further inform current SLD research and pedagogy. In fact, in a case study conducted by Hardy, Römer, and Roberson (2015), they observed the implementation of this corpus in a first-year composition classroom, which, according to student feedback at the end of the course, provided a wide variety of useful models in different disciplines and genres from which they could draw on to enhance their academic writing. This study demonstrates how MICUSP can provide models of expert-level writing that instructors may use to facilitate more authentic practice in the target register (Hardy, Römer, & Roberson, 2015). Further informing ESL/EFL pedagogy, corpus studies have also observed lexical bundles and phrase-frames in expert writing to identify key syntactic units which may be useful for writing pedagogy. Bychkovska and Lee's (2017) study on lexical bundles compared native-speaker argumentative essays from the MICUSP corpus and ESL texts from the Corpus of Ohio Learner and Teacher English (COLTE), finding that in both corpora, a tendency toward more phrasal styles of writing was indicative of highly rated writers who were developing increased proficiency in their compositions. Concerning PFs, several studies have been conducted on MICUSP. For instance, the creators of this corpus, Römer and O'Donnell (2009) have examined positional variation of 3-5 word PFs with only a variant in the medial slot in the works of native and non-native academic writers, finding that particular PFs occupy positions in paragraphs, sentences, and whole texts. O'Donnell, Römer, and Ellis (2013) also conducted a comparative analysis of corpora, still using the 3-5 word PF units, between MICUSP, LOCNESS, and ICLE, finding that manual analysis of PFs is necessary to fully recognize their semantic and discourse functions.

At the time of writing, no research has been conducted specifically examining variability, predictability, and function of PFs in the argumentative essay sub-corpus of native and non-native writers' works in MICUSP for the creation of genre-specific pedagogical materials. Therefore, this study seeks to provide insightful analysis of PFs identified in this sub-corpus which could then be applied toward the creation of pedagogical materials encouraging more focused SLD in argumentative essay composition.

III. OVERVIEW OF CURRENT STUDY

It should first be noted that in the context of the MICUSP corpus and for the purpose of this research study, there will not be a distinction where native speaker writing is considered the "better" or "ideal" alternative to non-native writing. Rather, the terminology of 'expert' will be used to recognize any writer within the MICUSP corpus who attained an 'A' grade on their composition (Flowerdew, 2017). This approach aligns with the study conducted by Römer and Arbor (2009) which asserts that in higher-level academic writing, writers tend to display more overlap in their language production so that nativeness is less important than expertise in a subject. This means that even non-native writings would be considered expert level compositions since they were designated the same high rating and, at this level, may display more similarities rather than differences in language use compared to native texts, though this assertion will be more carefully examined in this study through the proposed research question. Additionally, as addressed in the introduction, argumentative essays were the genre of choice for analysis in this study, as argumentative writing proves to be one of the more challenging genres for learners to grasp in their writing development.

In emulation of previous studies focusing on PFs in writing (Garner, 2016; Juknevičienė & Grabowski, 2018; and Grabowski, 2015), *kfNgram* was implemented as the (Fletcher, 2002-2012) text analysis tool to analyze most frequent 4-word PFs in this sub-corpus of native and non-native argumentative essays. These studies also guide the selection of

variability, predictability, and function as the key factors of investigation for PFs in this study in order to establish a phraseological profile of academic texts in this corpus while answering the following research question:

Do native and non-native expert academic writers display significant differences in syntactic complexity through their use of PFs considering:

- a. Variability?
- b. Predictability?
- c. Functional characteristics?

IV. METHOD

This section discusses the details of the corpus composition as well as the criteria for analysis used to help answer the research question of this study.

A. Composition of Corpus

The original sub-corpus of native and non-native argumentative essays selected for investigation in MICUSP consists of 186 papers total. Once the sub-corpus of 186 native and non-native argumentative essays was extracted from MICUSP in the form of PDF documents, all files were converted to .txt format. In this format, tags labeling the texts as part of the MICUSP corpus were manually removed. Originally, the corpus was composed of 27,393 word types and 561,013 tokens. These numbers were determined by removing the corpus tags and then running all 186 files through *AntConc 3.5.8* (Anthony, 2019), using the word list function. Due to the stark differences in sample sizes between the sub-corpora (for instance, there are far fewer words in the non-native sub-corpus than the native one), the data analysis approach used in Nekrasova-Beker's (2019) study was adopted. This method involves extracting equal sample sizes from the argumentative essays so that results are more comparable across corpora. The smallest corpus of non-native writer texts (78,088 tokens) was used as a basis of extraction for the corpus of native writer papers. As a result, 72 out of the 186 papers were analyzed. The following section discusses more extensively the extraction and filtration methods for PFs in this sub-corpus.

Native-speaker essays before extraction comprised much of the corpus (155 papers) while non-native speaker essays comprised the minority (31 papers). Therefore, the number of tokens for the non-native speaker essays (78,088) was used as the basis for extraction from the native corpus and all papers in the non-native-speaker corpus were unmodified (meaning no number of tokens were cut from the original papers) for analysis. The non-native speaker corpus had papers from the following disciplines: Biology, English, Natural Resources, Nursing, Philosophy, Political Science, Psychology, and Sociology. Therefore, the papers extracted from the native speaker corpus were also from these same disciplines so that only 41 papers were included. Tokens were extracted so that the number of tokens per discipline type in the native corpus matched those of the non-native corpus. The method of extraction was done in this way to make sure that there were an equal number of tokens in both corpora of analysis and to reduce the influence of discipline-specific language represented by drastically more tokens in the original native speaker data set.

B. Analysis

As addressed before, the program *kfNgram* (Fletcher, 2002-2012) was used to generate PFs for this study. First, n-grams of size 4 with a floor of 1 were generated, then using those same parameters, phrase-frames with four words and occurring at least once in the corpus were extracted. Four-word PFs were chosen as the unit of analysis because it has been used successfully by many previous PF studies such as those of Garner (2016) and Jukneviene & Grabowski (2018), since "...they have a more readily recognizable range of structures and functions than the shorter sequences and are more frequent than the longer ones" (Jukneviene & Grabowski, 2018, p. 307). PFs with variable slots in either the initial or final position (*BCD and ABC*) were included along with those in the medial position in the hopes of capturing the variation in usage of PFs for this corpus. First, any PFs that were only prevalent in one discipline type such as "in french polynesia *" and "toxic cane toads *" were removed; PFs had to occur in at least two different disciplines to be included for analysis since the goal of this study is to identify cross-disciplinary frames for native and non-native speakers. Then, frames that had numbers as part of their structure were excluded such as "tree ## ###-### *" and "and * ##### in." Instances where a number occupied the variable slot for a frame were also not counted toward final calculations and analysis for variants. Discipline-specific frames such as "the native species *" and topic-specific frames such as "moll is a *" were eliminated from the final analysis to identify frames that could be used regardless of discipline and topic. Nonsensical frames such as "* â œgreatâ and" and any 4-frames that did not serve a distinct pragmatic or discourse function such as "d by * and" and "s * it is" were also removed. Finally, the units were checked for correct spelling so that misspelled units were combined with their correctly spelled counterparts. Following the example of Lu, Yoon, and Kisselev (2018), with the data collected, a list of pedagogically useful frames was compiled as part of the sample teaching materials for this study. Only the top 60 most frequent phrase-frames were kept from the native and non-native speaker corpora so that there were a total of 120 frames to be put into functional and structural categories and analyzed for their variability and predictability. After combining phrases in common between the corpora, 103 total frames were identified and 15 of these frames were selected for a sample PF list in Appendix A.

C. Variability

In emulation of Garner's (2016) study, the same calculation method was employed to find out the difference in variability between groups. This was accomplished using a variant/p-frame ratio (VPR) for each 4-frame, showing how fixed or variable the structure is in its use (Römer, 2010). This was calculated by dividing the number of variants in a PF from the number of times it occurs in the corpus. For instance, the frame “* it would be” has seven potential variants (that, than, fact, seems, important, whatsoever, now) that could occupy that slot. In total, it occurs in the corpus ten times. Therefore, to calculate VPR for this frame, it would be seven divided by ten, yielding a value of 0.70. A VPR closer to one shows that the slot is highly variable, whereas a value closer to zero reveals that it is more fixed. Classifications for this study are the same ones used in the Garner (2016) study: fixed- $VPR < .30$, variable- VPR between $.30-.70$, and highly variable if $VPR > .70$. Based on these designations, the frame “* it would be” would then be classified as variable. Frames such as “on the * hand” (with a VPR of 0.12) which had few possible variants and a high occurrence in the corpus fell into the fixed category. Highly variable frames such as “does not * the” were typically characterized by a high number of possible variants and a high number of occurrences. So, in the case of this frame, it had six potential variants and occurred six times total in the corpus. This means that its VPR was one and with every occurrence it had in the corpus, there was a different variant occupying its slot.

D. Predictability

Again, using the same calculation method as Garner (2016) and Gray & Biber (2013), predictability was calculated by dividing the number of times the highest occurring variant occupied a slot by the total number of occurrences for that frame in the corpus. As an example, the variant *others* occurred the most within the frame “the rights of *” (five times) compared to other potential variants for that structure. The frame itself occurred within the corpus seven times. In dividing five by seven, the predictability value for that frame is 0.71. Frames were classified using the following system: highly unpredictable (< 0.25); unpredictable ($> 0.25 \leq 0.50$); predictable ($> 0.50 \leq 0.75$); and highly predictable (> 0.75). According to this system then, the example frame is predictable. Highly unpredictable frames such as “would like to *” (with a predictability value of 0.23) were often characterized by their highest occurring variant having a low value, while the overall number of occurrences for the frame in the corpus were high. Unpredictable frames like “it would be *” (with a predictability value of 0.50) had a moderate number of occurrences for their most common variant in comparison to the overall number of occurrences for that frame in the corpus. Lastly, predictable and highly predictable frames had high values for their most occurring variant that was closer to the number of occurrences of that frame in the corpus. For instance, the frame “on the * hand” had *other* as its highest occurring variant (it occurred ten times out of the 17 times the frame was used in the corpus), so that with a predictability value of 0.59, it would be a predictable structure writers could use.

E. Function

Functional categories were determined by manual analysis based on Grabowski's (2015) approach which stresses that PFs can be sorted into distinct categories independently from the semantic manifestations of their slot fillers in a text. This study utilizes the same classification system as Biber, Conrad, & Cortes (2004) in considering the function of 4-frames. The three possible classifications are referential expressions (like “theory of the *” or “the level of *”), stance expressions (including frames like “it is not *” and “can be * as”), and discourse-organizing expressions (such as “in order to *” and “on the * hand”). Within the context of this study, referential frames are used to describe or identify attributes in a text. Stance frames show the writer's disposition toward a particular subject/idea in their composition. And, finally, discourse-organizing frames reveal relationships between different pieces of language or ideas within the paper. In addition to these primary level classifications, the PFs were analyzed using secondary level classifications as devised by Simpson-Vlach & Ellis (2010) and Biber et al. (2004) to further refine the functions of PFs in this corpus. This is in emulation of the Lu et al. (2018) study which also used this method of classification in determining functions of PFs in introductions. Therefore, the second-level categories that were used to refine functional classification are the following:

- a. Referential Expressions: Specification of Attributes, Identification and Focus, Contrast and Comparison, Deictics and Locatives, & Vagueness Markers.
- b. Stance Expressions: Hedges, Epistemic, Ability, Evaluation, Obligation, and Intention.
- c. Discourse-Organizing Expressions: Metadiscourse and Textual Reference, Topic Introduction and Focus, Topic Elaboration, and Discourse Markers

V. RESULTS AND DISCUSSION

This study's aim was to determine if there were significant differences in phrase-frame usage among expert writers in argumentative essays from MICUSP concerning variability, predictability, and function. Detailed below are the results of this inquiry along with a discussion in each section of how they compare to the results of previous PF studies investigating these factors.

A. Variability

Out of 60 frames in the native corpus, none were identified as fixed, eight were identified as variable (13%) and 52 were identified as highly variable accounting for 87% of the corpus. In the non-native writer corpus, there was only one fixed frame (2%), 15 variable frames (25%), and 44 highly variable frames which make up 73% of the corpus. Based on these descriptive statistics, it was determined that concerning variability, both sets of data did not have a normal distribution; therefore, the Mann Whitney U test was implemented as a non-parametric statistical measure to determine if the differences in the data were statistically significant. Significance level was set at 0.50 with a 2-tailed hypothesis, with a p-value less than 0.05 indicating that the differences are indeed significant. Upon running this analysis between these data sets, the p-value was 0.00008, showing that variability between native and non-native expert texts is statistically significant. These results could be attributed to the drastic trends in the native-speaker corpus where there are no fixed frames and a high amount of highly variable frames. Such results indicate that perhaps native writers are more experimental in their use of the many possible variants that could occupy a frame compared to their non-native counterparts. Additionally, these native writers could have had more exposure to and practice with highly variable frames through their coursework.

While these results confirm that native writers generally use more highly variable frames, there is also a general trend in both groups of expert writers to use highly variable frames. These results make sense since, according to the selection method for frames, it did not exclude function-word frames such as *the * of the*, and as supported by Grabowski's (2015) study concerning PFs in pharmaceutical texts, function-based frames present the most phraseological variation. He found this especially the case with academic pharmaceutical discourse so, perhaps, the same could apply to other brands of academic discourse from various disciplines. Nekrasova-Beker (2019) also confirmed that in her research on the use of PFs in an engineering corpus, there were frames in common with other types of academic texts, suggesting that these function-based frames could be more indicative of the register of academic prose. Even between this study and Nekrasova-Beker's (2019) the function based frame "on the * of" is shared indicating a thread of commonality in the frames used for high-level academic writing. Thus, because this corpus's identified PFs are primarily function word based, they present a great deal of variability in the works of these expert writers, confirming what Garner asserted in his study (2016) that at higher levels, writers present compositions with more variable structures.

B. Predictability

Distributions regarding predictability in both corpora were also not normal, again justifying the application of the Mann Whitney U test to determine statistical significance. Parameters for this test were the same as those used to measure variability with a significance level of 0.50 and a 2-tailed hypothesis. The p-value was 0.65, indicating that the differences in predictability between the groups were not statistically significant. This means that overall, expert writers are consistent in their usage of mostly highly unpredictable and unpredictable frames in their writing, which again aligns well with Garner's (2016) general observation that higher level writing becomes less predictable in frame usage.

Again, perhaps due to the selection process of frames for this study, as well as the selection of frames across several disciplines, many of them were identified in both corpora as predominantly unpredictable or highly unpredictable. This could be because of the inclusion of function-based frames which allow more variation in their respective slot. For instance, a function-based frame like "on the * of" allows for as many as 35 different kinds of variants in the corpus. In contrast, verb-based frames such as "does not * the" (with 6 potential variants in the corpus) and content word-based frames like "the rights of *" (with 3 potential variants) often allow for less variation when filling their slot. Such results contrast those attained by Garner in his study (2016) which identified more highly predictable frames. However, such results could confirm that in addition to more common PFs across disciplines being highly variable, the variants themselves are also highly unpredictable (Nekrasova-Beker, 2019). This would again support the idea that with greater inclusion of other disciplines within a corpus of the same genre, the unpredictability of variants increases, since there are more possibilities for the kinds of lexical items that could fill said slot. In looking at academic prose as a register, this feature of higher unpredictability would then make sense and, perhaps, more discipline-specific analyses of PFs within a register would be needed in addition to the studies that have thus far been conducted to determine what more predictable frames and variants look like according to discipline and topic (Cunningham, 2017; Grabowski, 2015; Lu et al., 2018; Nekrasova-Beker, 2019).

C. Function

Of the functional categories, referential PFs were the most prevalent in the corpus, stance PFs were second most frequent, and discourse PFs were the least frequent (see Table 1 below). Conducting a Chi-square test for this data, there was also no statistically significant difference in usage of functional frames between the two groups of writers ($p=0.82$). This could indicate that generally high level writers within this corpus have a tendency toward using a high number of referential and stance frames regardless of their status as a native or non-native writer. So then for this section of analysis, both corpora were analyzed together to understand the common trends of expert writers concerning frame function. Comparing the overall results to the Lu et al. (2018) and the Garner (2016) studies which also employ the same Simpson-Vlach (2010) classification system, these results are to be expected since in both studies, referential PFs either make up a great proportion of usage in the corpus, or are most frequent overall, respectively. Therefore, the results of this study suggest that referential PFs play a prominent role in expert writing of academic prose.

TABLE I
DISTRIBUTION OF FUNCTIONAL FRAMES IN NATIVE AND NON-NATIVE WRITER CORPORA

Writer Status	Referential PFs	Stance PFs	Discourse PFs	Grand Total Frames
Native Speaker	38	19	3	60
Non-Native Speaker	34	19	7	60
Total	72	38	10	120

In the rest of this section, the sub-types of functional PFs are briefly discussed with a couple of examples of these frames being used for their designated function from the corpus.

Referential PFs make up most of the corpus. Concerning sub-types, *specification of attributes* were the highest occurring PFs within this category while *identification and focus* frames were second most occurring. No *deictics and locatives* (PFs indicating temporal, spatial, or environmental reference points) were identified in the corpus (Simpson-Vlach & Ellis, 2010). This indicates then that within this corpus of expert texts, writers prioritized specifying characteristics of concepts within their writing and introducing ideas or focal points for their composition. These results align with those of the Lu et al. study (2018) which also identified these sub-types as prominent features in social science research articles.

a. Referential PFs

- i. Specification of Attributes²(a PF identifying specific characteristics of a following nominal or clause):

1. Example: *A study on the [effects] of introduced trout on frogs showed that native populations can dramatically rebound upon extirpation of the invasive predator (Vredenburg 2004).*

From a Biology text written by a non-native senior undergraduate writer.

- ii. Identification and Focus(PFs that introduce an idea or focus of a paper/section):

1. Example: *It is [said] that Arendt has always refused to be labeled as a philosopher since she thought philosophy is concerned with man in the singular...*

From a Sociology text written by a non-native 1st year graduate writer.

Of the phrase-frames identified as Stance PFs, *evaluation* PFs were most common, with *epistemic* frames being the second most common type of frame. No frames concerning *obligation* (PFs that ask readers/listeners to perform a certain action) were found (Simpson-Vlach & Ellis, 2010). Based on this data then, it seems that when composing argumentative essays, statements of evaluation and expressions of personal views/opinions towards an idea take priority to create a successful piece of writing. Something worth noting is that *obligation* frames were identified in the Lu et al. (2018) study though it may be due to difference in genre since the focus of their study was to observe PFs in social science research article introductions whereas this study observed prominent PFs in argumentative essays. Another consideration is that this difference could be due to discipline specific practices where social science articles exhibit this type of PF specifically, whereas other disciplines may not.

a. Stance PFs

- i. Evaluation (PFs that evaluate a claim/idea in the text):

1. Example: *James' beliefs which result from memory processes should not be [counted] as justified.*

From a Philosophy text written by a native senior undergraduate writer.

- ii. Epistemic (PFs that function as personal knowledge claims or reports of others' knowledge on a topic):

2. Example: *It is [clear] that he sees nothing more perfect in life than a happy, loyal family.*

From an English text written by a native senior undergraduate writer.

Most frequent types of PFs within the Discourse-Organizing category were *topic elaboration cause and effect* and *discourse markers*. No *metadiscourse and textual reference* frames (indicating the outline/organization of the document) were identified. Prevalence of those two categories again aligns with the Lu et al. (2018) piece demonstrating how expert writers will employ PFs to show cause and effect relationships as well as connections between content. Lack of *metadiscourse and textual reference* PFs in comparison to the Lu et al. (2018) study could again be attributed to a difference in genre or discipline. Perhaps, at least in this corpus of argumentative essays, such PFs are not needed.

a. Discourse-Organizing PFs

- i. Elaboration Cause and Effect (frames indicating a cause-effect relationship)(Simpson-Vlach & Ellis, 2010):

² All definitions for functional sub-categories come from Lu et al. (2018) unless otherwise stated.

2. Example: *This is a form of self-handicapping, which occurs when people take actions to hurt their own performance or image in order to [protect] their own self-esteem (Berglas & Jones, 1978).*

From a Psychology text written by a non-native senior undergraduate writer.

- ii. Discourse Markers (PFs which act as connectives signaling transitions in a text) (Simpson-Vlach & Ellis, 2010):

1. Example: *The conscription system, on the [other] hand, exploits physical resources from the morally virtuous citizen.*

From a Sociology text written by a non-native 3rd year graduate writer.

VI. CONCLUSION: LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH AND PEDAGOGY

This study's aim was to investigate common trends in PF usage among native and non-native expert writers from the argumentative essay portion of the MICUSP corpus. These common trends were investigated by analyzing predictability, variability, and function of the most common frames used by these writers. Four-word frames were selected for study including function word-based PFs with the intent to create pedagogical materials based on the results of this study. While the corpus used for this study was relatively small (78,088 tokens for each collection of native and non-native speaker texts totaling 156,176 tokens total for the entire corpus), the study itself utilized some already well-established methods from the field of ESL/EFL research, such as the selection of 4-word PFs (Garner 2016), calculation of variability and predictability (Garner, 2016), inclusion of function word-based PFs (Grabowski, 2015), and functional (Biber et al., 2004; Simpson-Vlach & Ellis, 2010) PF classification systems. Through these methods, some PFs were identified and extracted from the corpus which could prove pedagogically useful (refer to Appendix A).

It should be acknowledged, however, that this study had limitations in terms of time constraints and some issues which naturally come from using the MICUSP corpus. As addressed in the work of Bahrami, Dowlatabadi, Yazdani, & Amerian, (2018), there is a need for understanding how expert writers use stance and other metadiscourse markers in their writing. For this study, Grabowski's (2018) fixed-frame-based approach was primarily used to address this need and categorize frames for their function. Categorization was done this way due to time constraints. While this is certainly a useful approach which focuses on structural rather than semantic characteristics of the frames, it could overlook multifunctional frames which may be more easily identified through a more variant-based approach. This may be an interesting venue to pursue in future research to further differentiate how expert writers employ frames for various uses in argumentative essays. In addition to implementing this strategy for analysis, it would also be helpful in a future iteration of this study to have another coder for functional designations of frames. This would ensure further reliability of the classifications for those structures. Another issue encountered in the construction of the study was the size of the corpus. Though this was somewhat managed by choosing comparable sample sizes, it would be useful to redesign the study so it's analyzing a greater data set across more disciplines, ensuring that the trends observed in this study could be used to inform cross-disciplinary writing instruction for argumentative essays.

Other corpora could also be examined using the measures of predictability, variability, structure and function. For instance, in emulation of Jukneviene & Grabowski's (2018) study, influence of L1 background could be more diligently analyzed among expert and developing student writers. In their study, they focused on Lithuanian and Polish speakers in the ICLE, though this could be done with other L1 backgrounds represented in that corpus. Such investigation would be especially timely as the third version of the ICLE is released for the ESL/EFL research community's use to identify trends among different L1 writers. Besides L1 backgrounds, expert writing could be analyzed even within the MICUSP corpus based on student levels so that a future study could analyze key differences in PF usage between senior undergraduates, and 1st-3rd year graduates to discern more subtle differences/potential trends in development between expert writers at the university level.

As asserted by previous studies in EFL research and pedagogy, explicit instruction in referential, stance and discourse markers is needed to demonstrate how to create successful academic texts (Candarli, Bayyurt, & Marti, 2015; Crosthwaite & Jiang, 2017; Escobar & Fernandez, 2017; Jomaa & Alia, 2019; Lee & Deakin, 2016; MacArthur, Jennings, & Philippakos, 2019; Manan & Raslee, 2018; Povolna, 2012). For instance, Crosthwaite & Jiang, (2017) found in their longitudinal study that students incorporated more hedges and less boosters and self-mentions (with an overall decrease in stance markers) leading to higher rated argumentative essays as a result of EAP instruction over time. Lee & Deakin (2016) also identified a greater use in hedges as a feature common to successful academic texts for undergraduate writers. Referential cohesion also contributed to more effective compositions for college level writers as well, as presented in the work of MacArthur, Jennings, & Philippakos (2019). Finally, discourse markers have been established as an especially crucial element for the higher levels of academic writing to help enforce coherence and cohesion (Povolna, 2012). More recent studies such as those of Candarli, Bayyurt, & Marti (2015) and Jomaa & Alia (2019) also assert that exposure to corpora of texts displaying target usage of these functional devices would prove beneficial to developing academic writers. Such exposure would allow them for opportunities to identify, discuss, and implement these features in their own writing. When designing pedagogical materials then, it may be a worthwhile task to include a healthy mix of PFs that function as referential markers since they are most frequent in academic writing, as well as discourse and stance markers (as displayed by the sample list in Appendix A). This may prove useful in creating

an environment where learners could hone their writing abilities, developing the right amount of explicitness, elaboration and complexity that is expected in the register through these functional frames. Lists like this would need to be rated for pedagogical usefulness by instructors and students alike, following the example set by Lu, Yoon, and Kisselev (2018). In addition, to make them more reliable as a pedagogical tool, further studies should be conducted to observe the relationship between student retention/application of PFs when using PF lists in conjunction with other class materials. And while studies such as those of Barabadi, Robotjazi, & Bayat (2020) and Golparvar & Barabadi (2020) stress the importance of developing and implementing PF lists for instruction, they do not always address how to do so. This study proposes combining PF lists with other available resources such as corpora/other target texts for academic writing engaging students in activities that demand first recognition and then practice and successful application of these structures. Appendix B offers a sample recognition activity that could be adapted for use in an EAP classroom with a target text based on the compiled list from this study.

The results of this study also contribute to the ever-growing body of research which intends to situate PFs as perhaps a more sensitive measure to detect syntactic complexity in academic genres. Differences in the use of variable frames among native and non-native expert writers could suggest that in alignment with other studies, such as Garner's (2016), variability is a defining element of successful academic writing and that as writers become more familiar with potential variants for a frame, they are more likely to implement them in their writing. Additionally, when examining predictability between the groups, it appears that both groups of writers use less predictable frames in their writing. Similar tendencies in frame usage were also found concerning function. Thus, based on this study, the differences in PF usage among expert writers is slight (concerning variability) and best measured through PFs. This should, however, be further examined by conducting more corpus studies with MICUSP and other sets of L2 writing data. In addition to their role as a potentially more sensitive measure for syntactic complexity, they present a pedagogical opportunity to lessen the cognitive strain experienced by ELLs.

Many studies still indicate that teaching language in processable chunks, such as lexical bundles, or potentially phrase-frames, is beneficial in ESL/EFL instruction to foster writing development for various rhetorical contexts (Araghi, Oskuee, & Salehpour, 2014; Chen & Baker, 2010; Hong & Hua, 2018; Hyland, 2012; Kazemi, Katiraei, & Rasekh, 2014; Rashti & Mohammadi, 2017). For a long time, lexical bundles have been considered a useful language unit for both pedagogy and research. However, as with any language unit, there are some pedagogical issues in using that unit for language instruction. For instance, learners still have issues implementing them in their writing even after overt instruction. Rashti & Mohammadi (2017) in fact highlighted how learners may tend to overuse or misuse discourse organizers and prepositional phrases such as "in addition to," "furthermore," "in other words," and "as a result of" (p. 212). Additionally, when introducing lexical bundles to learners, they may be asked more often to memorize chunks such as "on the other hand" and "on the one hand" as separate expressions, rather than as a single frame with other potential slot fillers. Therefore, the advantage of using phrase-frames as a unit of instruction may be that they demand learners process language inductively, accounting for linguistic variation in writing, rather than learning through rote memorization. PFs may then be the answer to enabling less strenuous cognitive processing in academic writing with the implementation of pedagogical materials towards this aim.

APPENDIX A. POTENTIALLY USEFUL PHRASE-FRAMES FOR ACADEMIC WRITING

Helpful Phrases for Academic Writing

Please keep this handout to help guide you while you write your essay. We will refer to these phrases often to enhance your writing.

Phrase-Frame	Slot Fillers	# of Potential Fillers	Functional Category
1. on the * hand	other (10), one (7).	2	Discourse Frame; Discourse Marker
2. out of the *	question(3), ordinary, material	3	Stance Frame; Evaluation
3. the rights of *	others(5), women, individuals	3	Referential Frame; Specification of Attributes
4. no * of the	knowledge, culture, fact (5), interpretation, facts.	5	Stance Frame; Evaluation
5. is that * is	there (4), it (2).	2	Stance Frame; Evaluation
6. * and the society	individual (4), individuals.	2	Referential Frame; Contrast & Comparison
7. the same * of	level (2), amount, way, understanding.	4	Referential Frame; Contrast & Comparison
8. *the relationship between	is (2), with, redefine, discussing.	4	Referential Frame; Contrast & Comparison
9. * it seems that	then(2), however, indeed, where, later, clearly	6	Stance Frame; Hedge
10. would be * to	committed, best, unlikely (2), unable, uncomfortable, natural, worthwhile, important, impossible, able, required, prone, difficult.	13	Stance Frame; Epistemic
11. should not be *	permitted (2), taken (2), made, overlooked, described, decided, articulated, justified, counted, removed, manipulated, excused.	12	Stance Frame; Epistemic
12. can be * as	interpreted (2), categorized, rewritten, restated, regarded.	5	Stance Frame; Epistemic
13. it is * to	worthy, useful, preferable, desirable, important (5), mistaken, difficult (2), enough, likely, ideal, up, hoped.	12	Stance Frame; Evaluation
14. in order to*	prevent (5), punish (2), provide (2), make (2), protect (2), maintain (2), determine (3), get (2), know (4).	46	Discourse Frame; Topic Elaboration Cause & Effect
15. to * with the	deal (6), cope (2), live, communicate, do (2), compete.	6	Discourse Frame; Discourse Marker

APPENDIX B. SAMPLE ACTIVITY COMBINING CORPUS WORK WITH PHRASE-FRAMES

Checklist for Phrase-Frames

Directions: Please use this checklist to identify the target frames in the text we have been working with today.

1. Underline all target frames that you can find in the text provided. Keep in mind, not all of them from our list may be found in the text.
2. Then, using the checklist, determine which frames have been used, and their function in the text (you may use your handout of frames to help you).

Frames	Is it included in the text?	If so, where is it in the text? (Include at least one example and the page number/s)	Variants used in the frame	Function(s) in the text
1. on the * hand	<input type="checkbox"/> Yes <input type="checkbox"/> No			
2. out of the *	<input type="checkbox"/> Yes <input type="checkbox"/> No			
3. the rights of *	<input type="checkbox"/> Yes <input type="checkbox"/> No			
4. no * of the	<input type="checkbox"/> Yes <input type="checkbox"/> No			
5. is that * is	<input type="checkbox"/> Yes <input type="checkbox"/> No			
6. * and the society	<input type="checkbox"/> Yes <input type="checkbox"/> No			
7. the same * of	<input type="checkbox"/> Yes <input type="checkbox"/> No			
8. *the relationship between	<input type="checkbox"/> Yes <input type="checkbox"/> No			
9. * it seems that	<input type="checkbox"/> Yes <input type="checkbox"/> No			
10. would be * to	<input type="checkbox"/> Yes <input type="checkbox"/> No			

ACKNOWLEDGMENTS

The author wishes to thank Drs. Tatiana Nekrasova-Beker and Tony Becker for their insightful expertise and guidance for this study.

REFERENCES

- [1] Anthony, L. (2017). Antconc (Version 3.5.8) [computer Software]. Tokyo, Japan: Waseda University. Available from: <http://www.laurenceanthony.net/software> (accessed 01/09/2019).
- [2] Araghi, S.M., Oskuee, M.Y., & Salehpour, S. (2014). The Effect of Pre-teaching Extended Prefabricated Lexical Bundles on the Writing Development of Advanced EFL Learners. *International Journal of Applied Linguistics & English Literature*, 3.1, 214-219.
- [3] Bahrami, L., Dowlatabadi, H. R., Yazdani, H., & Amerian, M. (2018). Authorial Stance in Academic Writing: Issues and Implications for Research in English Language Teaching. *International Journal of English Language & Translation Studies*, 6.2, 69-80.
- [4] Barabadi, E., Robotjazi, M. A., & Bayat, M. (2020). A phraseological examination of research Articles in the field of environment using key phrase frame. *Eurasian Journal of Applied Linguistics*, 6.1, 81-100.
- [5] Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25.3, 371-405.
- [6] Biber, D. & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, Elaboration, explicitness. *Journal of English for Academic Purposes*, 9, 2-20.
- [7] Biber, D., Gray, B., & Ponpoon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45, 5-35.

- [8] Bychkovska, T. & Lee, J.J. (2017). At the same time: Lexical bundles in L1 and L2 university Student argumentative writing. *Journal of English for Academic Purposes*, 30, 38-52.
- [9] Candarli, D., Bayyurt, Y. & Marti, L. (2015). Authorial presence in L1 and L2 novice academic writing: Cross-linguistic and cross-cultural perspectives. *Journal of English for Academic Purposes*, 20, 192-202.
- [10] Chen, Y.-H. & Baker, P. (2010). Lexical Bundles in L1 and L2 Academic Writing. *Language Learning and Technology*, 14.2, 30-49.
- [11] Crosthwaite, P. & Jiang, K. (2017). Does EAP affect written L2 academic stance? A longitudinal learner corpus study. *System*, 69, 92-107.
- [12] Cunningham, K.J. (2017). A phraseological exploration of recent mathematics research articles through key phrase frames. *Journal of English for Academic Purposes*, 25, 71-83.
- [13] Escobar, C. F. & Fernandez, L. C. (2017). EFL Learners' Development of Voice In Academic Writing: Lexical bundles, Boosters/Hedges and Stance-taking Strategies. *Gist Education and Learning Research Journal*, 15, 96-124.
- [14] Fletcher, W. (2002-2012). KfNgram. A free software program. Available at: <http://www.kwicfinder.com/kfNgram/kfNgramHelp.html> (accessed 01/09/2019).
- [15] Flowerdew, J. (2017). Corpus-based approaches to language description for specialized academic writing. *Language Teaching*, 50.1, 90-106.
- [16] Garner, J.R. (2016). A phrase-frame approach to investigating phraseology in learner writing across proficiency levels. *International Journal of Learner Corpus Research*, 2.1, 31-68.
- [17] Golparvar, S. E. & Barabadi, E. (2020). Key phrase frames in the discussion section of research articles of higher education. *Lingua*, 236, 1-15.
- [18] Grabowski, L. (2015). Phrase frames in English pharmaceutical discourse: A corpus-driven study of intra-disciplinary register variation. *Research in Language*, 13.3, 266-291.
- [19] Gray, B. & Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics* 18.1, 109-135.
- [20] Hardy, J.A., Römer, U. & Roberson, A. (2015). The power of relevant models: Using a corpus of student writing to introduce disciplinary practices in a first year composition course. *Across the Disciplines*, 12.1, 1-20.
- [21] Hong, A.L. & Hua, T.K. (2018). Specificity in English for Academic Purposes (EAP): A Corpus Analysis of Lexical Bundles in Academic Writing. *The Southeast Asian Journal of English Language Studies*, 24.2, 82-94.
- [22] Hyland, K. (2012). Bundles in Academic Discourse. *Annual Review of Applied Linguistics*, 32, 150-169.
- [23] Jomaa, N. J. & Alia, M. M. (2019). Functional Analyses of Metadiscourse Markers in L2 Students' Academic Writing. *Arab World English Journal*, 10.1, 361-381.
- [24] Juknėvičienė, R. and Grabowski, L. (2018). Comparing formulaicity of learner writing through phrase-frames: A corpus-driven study of Lithuanian and Polish EFL student writing. *Research in Language*, 16.3, 303-323.
- [25] Kaewpet, C. (2018). Quality of argumentation models. *Theory and Practice in Language Studies*, 8.9, 1105-1113.
- [26] Kazemi, M., Katiraei, S. & Rasekh, A. E. (2014). The Impact of Teaching Lexical Bundles on Improving Iranian EFL Students' Writing Skill. *Procedia - Social and Behavioral Sciences*, 98, 864-869.
- [27] Kyle, K. & Crossley, S.A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, 102, 333-349.
- [28] Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30.4, 579-589.
- [29] Lee, J.J. & Deakin, L. (2016). Interactions in L1 and L2 undergraduate student writing: Interactional metadiscourse in successful and less-successful argumentative essays. *Journal of Second Language Writing*, 33, 21-34.
- [30] Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45, 36-62.
- [31] Lu, X. & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, 29, 16-27.
- [32] Lu, X., Yoon, J., and Kisselev, O. (2018). A phrase-frame list for social science research article introductions. *Journal of English for Academic Purposes*, 36, 76-85.
- [33] MacArthur, C. A., Jennings, A., & Philippakos, Z. A. (2019). Which linguistic features predict quality of argumentative writing for college basic writers, and how do those features change with instruction? *Reading and Writing*, 32.6, 1553-1574.
- [34] Manan, N. A. A. & Raslee, N. N. (2018). Explicit Discourse Marker Instruction to Improve Coherence and Cohesion in Academic Writing. *International Journal of Academic Research in Business & Social Sciences*, 8.1, 465-483.
- [35] Martinez, A. C. L. (2017). Syntactic complexity in secondary-level English writing: Differences among writers enrolled on bilingual and non-bilingual programmes. *Porta Linguarum*, 28, 67-80.
- [36] Martinez, A.C.L. (2018). Analysis of syntactic complexity in secondary education: EFL writers at different proficiency levels. *Assessing Writing*, 35, 1-11.
- [37] Michigan Corpus of Upper-Level Student Papers. (2009). *Distribution Across Disciplines* [Data file]. Retrieved from: <http://elcicorpora.info/> (accessed 01/09/2019).
- [38] Nekrasova-Beker, T.M. (2019). Discipline-specific use of language patterns in engineering: A comparison of published pedagogical materials. *Journal of English for Academic Purposes* 41, 1-12.
- [39] O'Donnell, M. B., Römer, U. and Ellis, N.C. (2013). The Development of Formulaic Sequences in First and Second Language Writing. Investigating Effects of Frequency, Association, and Native Norm. *International Journal of Corpus Linguistics* 18.1, 83-108.
- [40] Povolna, R. (2012). Causal and Contrastive Discourse Markers in Novice Academic Writing. *Brno Studies in English*, 38.2, 131-148.
- [41] Römer, U. (2010). Establishing the phraseological profile of a text type: The construction of meaning in academic book reviews. *English Text Constructions*, 3.1, 95-119.

- [42] Römer, U. & Arbor, A. (2009). English in Academia: Does Nativeness Matter? *Anglistik: International Journal of English Studies*, 20.2, 89-100.
- [43] Römer, U., & O'Donnell, M.B. (2009). Positional variation of phrase frames in a new corpus of proficient student writing. [Online] Paper presented at AACL conference. Edmonton, Canada, 9 Oct 2009.
- [44] Römer, U., & O'Donnell, M. B. (2011). From student hard drive to web corpus (part 1): the design, compilation and genre classification of the Michigan Corpus of Upper-Level Student Papers (MICUSP). *Corpora*, 6.2, 159–177.
- [45] Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31, 487-512.
- [46] Vercellotti, M. L. (2017). The development of complexity, accuracy, and fluency in second language performance: A longitudinal study. *Applied Linguistics*, 38, (1), 90-111.
- [47] Yang, W. & Sun, Y. (2015). Dynamic development of complexity, accuracy and fluency in multilingual learners' L1, L2, and L3 writing. *Theory and Practice in Language Studies*, 5.2, 298-308.
- [48] Yoon, H.-J. (2017). Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality. *System*, 66, 130-141.

Kseniya N. Walcott was born in Tyumen, Russia in 1995. Ms. Walcott attained her Bachelor of Arts in English with a language concentration and her Master of Arts in English with a focus on teaching English as a second/foreign language at Colorado State University in Fort Collins, CO, USA in 2017 and 2019 respectively.

She has worked as a Composition Instructor, Language Tutor, Research Assistant and College Test Proctor. She currently works as an Admissions Processor for Colorado State University and Private Language Tutor in Fort Collins, CO, USA. Some of her research interests include topics in neurolinguistics and novel ways to teach syntax to language learners.