Konin Language Studies
Faculty of Humanities and Social Sciences, State University of Applied Sciences in Konin, Poland
KSJ 9 (1). 2021. 11-34
http://ksj.pwsz.konin.edu.pl
doi: 10.30438/ksj.2021.9.1.1

Predicting scores in English argumentative essays by L1 Spanish and English authors: A broad exploration of writing quality indices

Joe Geluso
Nihon University College of Law, Japan
https://orcid.org/0000-0002-7496-481X
joe.geluso@nihon-u.ac.jp

Abstract

In recent years numerous studies have investigated how different writing features, such as lexical complexity and syntactic complexity, impact writing quality scores. However, most studies focus on only one set of features (e.g., lexical sophistication) to explain holistic writing quality scores. Thus, the purpose of the present paper is to use a broad set of features, covering lexical complexity, syntactic complexity, and cohesive properties, to explore which features were associated with and best predicted holistic scores of English essays among L1 Spanish and English authors. Findings from regression analyses suggest that lexical sophistication and phrasal complexity best predicted high scores in both the L1 Spanish and English groups. This finding supports previous research on lexical sophistication and syntactic/grammatical complexity (Biber et al., 2011; Kyle & Crossley, 2015). However, differences were found between L1 groups in the number of predictors, specific predictors, and amount of variance explained by each model. Furthermore, cohesion indices and multiword units played a limited role in predicting scores in this data set. The present study can provide teachers and learners with insights into the combination of features that predict higher writing scores from a broad set of features, and thus help to focus teaching and autonomous studies.

Keywords: writing quality; learner writing; natural language processing; corpus

1. Introduction

Research on writing quality has a long and storied history in the field of applied linguistics. Much research has aimed to understand the characteristics that predict higher writing quality scores. A clearer understanding of writing features that are likely to garner higher scores will be of pedagogical value as writing tasks are frequently used in a variety of contexts: from classroom assessment to placement tests to higher-stake tests such as TOEFL and IELTS used in determining university admissions. Research on writing features and writing quality scores has historically tended to focus on lexical sophistication, syntactic complexity, and cohesion.

Typically, writing quality studies focus on just one of these areas (i.e., lexical sophistication, syntactic complexity, or cohesion) as analytic writing rubrics used to score essays will separate these constructs between rating bands. Thus, a study on, say, lexical sophistication can focus specifically on the lexical aspects of essays in an effort to ascertain which characteristics of lexis are most likely to result in a high score within a rating band focused on lexis. However, scholars have called for studies that investigate a wider array of indices to predict holistic scores rather than relying on just one language proficiency area such as syntactic proficiency or cohesion (Kyle & Crossley, 2017; Yoon, 2018). This study aims to address this call by including measures that cover lexical, syntactic, and cohesive proficiency properties in an effort to predict holistic essay scores. The present study focuses on English essays authored by L1 Spanish and L1 English speakers.

2. Literature Review

2.1. Lexical complexity

Lexical complexity has been shown to be a strong predictor of writing quality, particularly in the form of lexical diversity and lexical sophistication (Bulté & Housen, 2014; Kyle & Crossley, 2015; Read, 2000). Measures of lexical diversity such as type-token ratio (TTR) can be used as a reflection of breadth of vocabulary knowledge, and TTR has been shown to be a powerful predictor of essay scores (González, 2017). TTR is calculated by dividing the number of word types (unique words) by the number of word tokens (total number of words). Type-token ratio measures range between 0 and 1. For example, the sentence *The ball bounced off the roof before hitting the ground* has 8 types and 10 tokens for a TTR of 0.8. In addition to measures of frequency and within-text lexical diversity, measures of range or dispersion across texts have received more attention as of late to measure usefulness of lexis in terms of breadth (Gries, 2008).

Lexical sophistication generally refers to the depth and breadth of lexical knowledge (Meara, 2005). Typically, corpus-based frequency counts are used to capture depth and breadth, comparing normalized frequencies from a set of target texts to normalized frequencies from a representative corpus or against frequency bands such as those in the Academic Word List (Coxhead, 2000; Laufer & Nation, 1995). Lexical sophistication studies have also considered the psycholinguistic properties of words based on human judgements of properties such as abstractness, word familiarity, and extent of word associations (Brysbaert et al. 2013; Kim et al., 2018; Kyle & Crossley, 2015). Words can also be analyzed in terms of their polysemy. Crossley et al. (2010) explain that polysemous words rest between homonymy and vagueness. A homonym, of course, is a word with two or more unrelated meanings such as *bat* as in a baseball bat or the winged animal. A vague word has only one meaning, such as *penicillin*. The word *class* is an example of a polysemous word as it can refer to related concepts such as "a class of people" in terms of social standing or "a class" of students in a school. Crossley et al. (2010) noted that polysemous words are more common than vague words and tend to be frequent. A final measure of lexical sophistication to be discussed here investigates the role of formulaic language or lexical bundles in writing (e.g., Biber et al., 2004; Cortes, 2004). Simpson-Vlach and Ellis (2010) created the Academic Formula List (AFL), a list of 3, 4, and 5-word sequences that are common in academic writing. The list can be used as a pedagogical tool for the teaching and learning of useful formulas for academic writing. This focus on multi-word units begins to blur the line with the next set of writing proficiency measures: syntactic and grammatical complexity.

2.2. Syntactic and grammatical complexity

Traditionally, research on complexity in writing has been grounded in ratio- and length-based measures that draw on the clause as the unit of analysis (see Ortega, 2003, for a review of classic studies on writing complexity). Examples include measures of the mean length of clause and number of dependent clauses per independent clause or per sentence. Research has found relatively small effect sizes between classic measures of syntactic complexity, which are predominantly clause-based, and holistic writing scores (Kyle & Crossley, 2017). This aligns with recent research findings that indicate that clausal complexity may better reflect complexity in speaking and phrasal embedding may better reflect complexity in writing (Biber et al., 2011; Ortega, 2003). Meanwhile, numerous recent studies have found that embedded phrasal structures, typical of informationally dense writing such as academic writing, become more prominent as writers become more proficient (Biber et al., 2011; Biber & Gray, 2010; Parkinson &

Musgrave, 2014; Staples et al., 2016). Phrasal embedding can include the simple embedding of adjective in noun phrases (e.g., *fast car*) to more complex embedding of prepositional phrases functioning as post nominal modifiers or adverbials (e.g., *an increase in broken homes; discuss this in the morning*).

In addition to phrasal embedding, Kyle and Crossley (2017) found that usage-based indices correlated more strongly with holistic writing scores than classic measures of syntactic complexity. Kyle and Crossley (2017) provide a detailed explanation of the various measures that are used to operationalize usage-based indices of syntactic sophistication, but by and large they rely on the analyses of verb argument constructions (VACs). Specifically, the association strength between VACs and the main verb lemma associated with a given VAC. For example, the verb *go* most frequently fills the verb slot in a verb locative (VL) construction such as *I go to the store* (Ellis & Ferreira-Junior, 2009).

2.3. Cohesion

Cohesion refers to the degree to which a text is connected based on textual features. Cohesive devices can function at local, global, and overall levels (Crossley et al., 2016). Local cohesion operates at the sentence level, global cohesion connects larger parts of a text such as paragraphs, and overall cohesion refers to an entire text but not specific connections between discrete chunks. Crossley et al. (2016) found that global indices of cohesion best predicted human judgements of cohesion in texts in comparison to local and overall cohesive indices. This finding seems in part to support Halliday and Hasan's (1976) seminal work, but it is surprising that overall cohesion indices were not found to be stronger predictors of human judgements. Halliday and Hasan viewed an individual text as a semantic unit, writing that "a text has texture . . . [and] it derives this texture from the fact that it functions as a unity[sic] with respect to its environment" (1976, p. 2). According to Halliday and Hasan (1976), texture is created by cohesive relations within a text, with different resources functioning to create texture (p. 2). They specifically focus on the following textual resources: (1) reference, (2) substitution and ellipsis, (3) conjunction, and (4) lexical cohesion.

Reference is generally achieved through pronouns used in anaphoric references to established actors in a text. Substitution and ellipsis are also typically anaphoric with the former substituting a new word or phrase (e.g., *mine*, *this one*) in reference to information established earlier in a text (e.g., *which book?*), and the latter requires the listener or reader to use given information to fill in an intentional blank. Conjunctions and/or connectives, refer to devices that connect preceding text to what follows (Halliday & Hasan, 1976). Examples of connectives include: *in sum*, *that is*, *to this end*, and *on the other hand*. Finally, lexical cohesion takes the form of lexical repetition and can be simple repetition of a word, complex repetition where derivational forms are counted as repetition, and more distant, implicit relations such as repeating themes with synonyms, or using relations of hypernymy or meronymy (Hoey, 1991). For example, the word *leaves* is a meronym of trees, so using *leaves* in a sentence following *trees* would be an instance of using meronymy for complex repetition.

The goal of the present study is to better understand what combination of indices from lexical complexity, syntactic complexity, and cohesion best account for higher holistic essay scores. The present study is exploratory in nature and draws on a range of natural language processing (NLP) tools to generate indices to represent these language proficiency areas. Specific research questions are:

1. Which indices of linguistic proficiency best predict holistic writing scores in English argumentative essays authored by L1 speakers of Spanish?
2. Which indices of linguistic proficiency best predict holistic writing scores in English argumentative essays authored by L1 speakers of English?
3. How do the indices identified in RQ1 and RQ2 differ between L1 speakers of Spanish and English?

3. Methodology

3.1. Corpus

Argumentative essays from the Spanish sub-corpus (ICLE-SP) of version 2 of the International Corpus of Learner English (ICLE; Granger et al., 2009) and the L1 English speaker counterpart corpus LOCNESS (Granger, 1998) comprised the corpora used in this study. Table 1 provides a general overview of these corpora. Word counts were generated from the NLP tools used for the analysis in this study, specifically the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC; Kyle, 2016) which is explained in more detail below. The texts were not originally annotated for topic, but topics are outlined in accompanying PDF files. Following Schanding and Pae (2018), each file was reviewed and categorized broadly into topical groups. Essay topics are presented in Appendix 1.

Table 1 Overview of ICLE-SP and LOCNESS sub-corpora

Corpus	Register	Author L1	Word count	No. of texts
ICLE-SP	English academic essays	Spanish	199,331	251
LOCNESS	English academic essays	English	258,165	297

3.2. Scoring essays

In order to assess writing quality, each essay was scored using a holistic rating scale adapted from the IELTS Task 2 Writing band descriptors (public version), accessed in April of 2018. The scale ranged from 1 to 7 with 1 being the lowest score and 7 the highest. The rating scale is shared in Appendix 2. Ratings were assigned by a group of 15 raters who were trained in either a face-to-face training session or via a telecommunication software such as Skype® or Google Hangouts®. The raters were graduate students in the English department of a large mid-western university in the United States and were paid to rate the essays. After completing the training sessions, each rater rated 126 or 127 essays within a two-week period. A Many Facet Rasch Measurement (MFRM) was used to generate an adjusted score or "fair score" for each essay (Eckes, 2015).

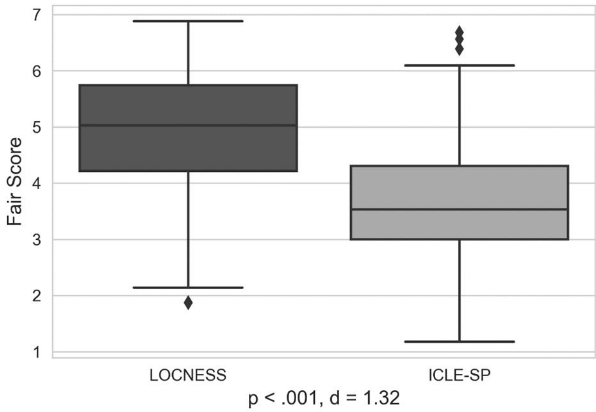

Figure 1 Boxplot representing fair scores between essays in LOCNESS and ICLE-SP

An MFRM generates fair scores by controlling for "rater severity." This is important because there will be variation in essay scores even between trained raters, and measures of consensus (e.g., Cohen's Kappa) or consistency (e.g., correlation) do not always imply high scoring accuracy (Eckes, 2015). A resulting fair score from an MFRM, however, reflects the score an essay would receive from a rater of average severity within a group of raters. Therefore, essays were rated by at least two, and sometimes three or four of 15 raters. Each rater overlapped with every other rater a minimum of eight and up to 15 times, resulting in a connected data set and allowed for the computation of fair scores. The MFRM was conducted using the free student version of FACETS called MINIFAC (Linacre, 2018), and the resulting fair scores were used to represent writing quality. Figure 1 displays the distribution of scores between the two groups. The mean essay score in LOCNESS was 4.97 with a standard deviation of 1.06. The mean essay score in ICLE-SP was 3.64 with a standard deviation of 0.95. As the assumption of equal variances was not met, a Welch's independent samples *t*-test was conducted between the ICLE-SP and LOCNESS essays. Fair scores between the two sets of essays were significantly different ($t_{(546)} = 15.45$, $p < .001$), with a large effect of $d = 1.32$.

In order to ascertain which writing measures best predicted fair scores within each group, two regression models were built: one for the ICLE-SP corpus, and the other for LOCNESS. Another possibility would have been to build a single regression model with L1 as a covariate. However, it was desired to see which combination of writing measures emerged as the best predictors of score without a different group of authors connected to the model.

3.3. Natural language processing tools

Recent developments in natural language processing (NLP) have provided methods for automating the analysis of large quantities of linguistic features. The tools used in the present study were developed by Kristopher Kyle and his colleagues and were used to extract a range of measures from different language proficiency areas including lexis, syntax, and cohesion. The three tools used to extract these measures include the Tool for the Automatic Analysis of LExical Sophistication (TAALES; Kyle & Crossley, 2015); the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC; Kyle, 2016); and the Tool for the Automatic Analysis of Cohesion 2.0 (TAACO; Crossley et al., 2016). In-depth explanations of the tools and the precise measures that they capture can be found in Kyle and colleagues' references above. Additionally, Kyle and Crossley (2018) provide a comprehensive explanation and useful examples of different indices generated by TAASSC. Readers are encouraged to refer to these publications to supplement the explanations of variables and output provided in the present study.

3.4. Variable selection

Regarding lexical diversity, TTR was calculated with a custom Python script that used a moving window approach. That is, for each essay, TTR was calculated on 400-word segments and then averaged for a final TTR measure. From there, all main options were activated in TAALES, which resulted in 254 indices of lexical sophistication. These measures included raw and transformed measures of frequency, range, and *n*-grams, and compared these measures to the Academic Word List (Coxhead, 2000) and subsections of the Corpus of Contemporary American English (COCA, Davies, 2008-). With regard to syntactic complexity, measures from TAASSC were used including 14 classic measures of syntactic complexity used in Lu (2011). In total, 368 measures were generated by TAASSC. Finally, for cohesive features, the default settings in TAACO were used to analyze the essays plus the Latent Semantic Analysis (LSA) feature for calculating within-text semantic similarity, for a total of 64 measures.

Once all indices were extracted, I followed the steps of other studies with similar goals (cf. Kyle & Crossley, 2015; Yoon, 2018). First, measures were checked for normal distributions using histograms. Those that were clearly not normally distributed were removed. Second, all measures were correlated with the overall fair scores of the essays to retain only statistically significant measures and those with a meaningful correlation with fair score. The correlation coefficient cutoff was set at $r \geq .20$. Finally, measures were checked for multicollinearity. Any two measures that had a correlation coefficient of .90 or higher were compared in terms of their correlation coefficient and *p*-value in reference to fair score, and the weaker measure was removed. For ICLE-SP, 639 measures were pruned from the initial 686 measures resulting in 47 remaining measures. For LOCNESS, 635 measures were pruned from the 689 initial measures resulting in 54 measures. To discover which combination of the remaining indices best predicted holistic essay scores in ICLE-SP and LOCNESS, the remaining measures were entered into two separate stepwise multiple regression models, one model for ICLE-SP and one for LOCNESS. The results of the correlation analyses and stepwise multiple regressions are presented in the next section.

4. Results

The correlation coefficients of all measures that correlated with fair scores at $r \geq .20$ are included in Appendix 3. This includes 47 predictors for ICLE-SP essays and 54 for LOCNESS. Appendix 3 also includes information on the type of measure (e.g., lexical complexity), along with the measure name.

To help with interpreting the regression models, Table 2 presents information about the relevant predictors for both ICLE-SP and LOCNESS. The category of each predictor (i.e., lexical, syntactic, cohesive) is presented, followed by the name of the predictor and a brief description of the linguistic construct that the predictor is intended to capture.

Table 2 Linguistic category and description of predictor variables in regression model

Category	Predictor	Intended linguistic construct
Lexical	Brysbaert_Concreteness_Combined_AW	Psycholinguistic measure of concreteness of words
	Brown_Freq_AW	Frequent words in Brown spoken corpus
	content_poly	Sum of polysemy scores for words in essay
	hyper_noun_S1_P1	Average hypernymy score for nouns (measure of semantic similarity)
	KF_Freq_AW	Frequent words in Brown written corpus
	Phono_N_H	Phonological neighbors
	SUBTLEXus_Range_FW_Log	Word frequency score in spoken language (subtitles from movies and TV series) from the SUBTLEXus corpus.

	USF_AW	Contextual distinctiveness/breadth (all words)
	USF_CW	Contextual distinctiveness/breadth (content words)
Syntactic	all_av_lemma_construction_freq_type	Average lemma construction frequency (types) - all COCA
	amod_dobj_deps_NN_struct	Adjectival modifiers per direct object (no pronouns)
	amod_pobj_deps_struct	Adjectival modifiers per direct object
	av_dobj_deps_NN	Dependents per direct object (no pronouns)
	news_av_lemma_construction_freq_log	Average lemma construction combination frequency, log Transformed - news
	news_av_lemma_construction_freq_type	Average lemma construction frequency (types only) - news
Cohesive	lsa_2_all_para	Relationship between concepts within a text

Tables 3 and 4 present the regression models for the ICLE-SP and LOCNESS essays, respectively. With regard to ICLE-SP, six measures accounted for 26.7% of the variance in predicting fair scores. Three variables were measures of lexical sophistication (entries 1, 5, and 6) and three of syntactic complexity (entries 2, 3, and 4).

Table 3 Stepwise multiple regression analysis to predict holistic essay scores in ICLE-SP

Entry	Predictor	*r*	R^2	R^2 change	B	SE	Beta
1	USF_CW	.291	.084	.084	-0.030	0.009	-0.203
2	news_av_lemma_construction_freq_log	.379	.144	.059	-0.536	0.234	-0.136
3	amod_dobj_deps_NN_struct	.426	.181	.038	1.528	0.475	0.181
4	news_av_lemma_construction_freq_type	.456	.208	.026	-4.24^{e-5}	0.000	-0.248
5	content_poly	.494	.244	.036	-.217	0.068	-0.191
6	SUBTLEXus_Range_ FW_Log	.517	.267	.024	-17.374	6.186	-0.160

Note. Estimated constant term = 76.178; B = unstandardized beta; SE = standard error; Beta = standardized beta

Table 4 presents the regression model with 10 measures that account for 41.4% of the variance in predicting fair scores in the essays in the LOCNESS corpus. Six measures comprised lexical sophistication: entries 1, 2, 6, 7, 8, and 9; three measures were of syntactic complexity: entries 4, 5, and 10; and one variable was a measure of cohesion, entry 3.

Table 4 Stepwise multiple regression analysis to predict holistic essay scores in LOCNESS

Entry	Predictor	*r*	R^2	R^2 change	B	SE	Beta
1	Brysbaert_Concreteness_Combined_AW	.285	.081	.081	-2.201	0.492	-0.246
2	hyper_noun_S1_P1	.398	.158	.077	0.776	0.160	0.245
3	lsa_2_all_para	.458	.210	.051	1.614	0.387	0.197
4	all_av_lemma_construction_freq_type	.498	.248	.038	-1.507^{e-5}	0.000	-0.329
5	amod_pobj_deps_struct	.557	.310	.062	2.388	0.662	0.203
6	Brown_Freq_AW	.587	.344	.035	0.004	0.001	0.583
7	KF_Freq_AW	.605	.366	.021	0.000	0.000	-0.490
8	Phono_N_H	.628	.395	.029	-0.130	0.036	-0.231
9	USF_AW	.636	.404	.009	0.039	0.016	0.142
10	av_dobj_deps_NN	.644	.414	.010	0.472	0.213	0.120

Note. Estimated constant term = 2.099; B = unstandardized beta; SE = standard error; Beta = standardized beta

5. Discussion

5.1. Correlation analysis

Correlations between fair score and various writing quality indices generated are too numerous to cover in fine detail here, so a select few are discussed to illustrate trends and help understand the array of variable names. As a reminder, variables and correlation coefficients are presented in Appendix 3. One striking result is the number of negative correlations with fair score in the lexical sophistication category in both ICLE-SP and LOCNESS texts. Many of these negative correlations can be explained by register differences. For example, the measures labeled SUBTLEXus are derived from a corpus of subtitles from television series and movies. Television and film dialogue resembles natural spoken English more than written academic English, and previous research by Biber and colleagues has shown clear differences in the linguistic characteristics between spoken and written registers (Biber & Conrad, 2009). Therefore, the negative correlation between SUBTLEXus and fair score is to be expected. Lastly, there were numerous variables with "KF" and "MRC" in the name (e.g., KF_Ncats_AW, MRC_Familiarity) in both lists that negatively correlated with fair score. Broadly, the former is related to range and the latter to familiarity. Words with high range scores are found across many texts and are likely more general words, and less context specific. Those words with higher familiarity scores are likely more frequent, and thus more familiar words. The negative correlation of these categories suggests that less frequent and more context specific words are correlated with higher fair scores.

While there are many negative correlations, there are also numerous measures that positively correlated with higher fair scores. For instance, in ICLE-SP, measures that one would expect to see are higher type token ratios (r = .242) and lexical density (r = .202). The former indicates that essays that demonstrated greater lexical diversity were generally rated higher, and the latter that more content words in essays were associated with higher scores. Likewise, the measure All_AWL_normed, referring to words from the Academic Word List, demonstrated a relatively strong correlation with fair score (r = .268), indicating that lower frequency, academic words were also associated with higher scoring essays.

Less common measures of lexical sophistication that positively correlated with fair score in ICLE-SP and LOCNESS included lexical decision reaction times (e.g., LD_Mean_RT_SD). In lexical decision time tasks, participants are shown a string of letters and must decide if the letters form a word or nonword. It is likely that longer reaction times are reflective of longer and more sophisticated words, hence the positive correlation with higher fair scores. The positive correlation here is congruent with Berger et al.'s (2019) finding that higher proficiency L2 writers tended to use words that elicited longer response times. This is likely the result of less frequent words requiring more time to identify and hence less frequent words correlating with higher scores on essays. Hypernymy was another measure that positively correlated with fair score, and hypernymic measures were better represented in LOCNESS than ICLE-SP. Hypernymy refers to superordinate and subordinate relationships between words. For instance, the term *automobile* is a hypernym, or superordinate, of *car* and *pick-up truck*, as the latter two are types of *automobiles.* Meanwhile, *car* and *pick-up truck* are the hyponyms, or subordinates, of *automobile*. Moving down a hypernymy chain will lead to more specific words. The fact that LOCNESS featured more and stronger correlations between hypernymy and fair score suggests more specific word choice among writers represented in LOCNESS.

With regard to syntactic measures, both corpora featured classic complexity measures from the Syntactic Complexity Analyzer (Lu, 2011) that positively correlated with fair score. These include clause-based measures such as mean length of clause (MLC) and complex nominals per clause (CN_C). The former measure has been found to be a significant, albeit weak, predictor of holistic writing scores in previous research (Kyle & Crossley, 2017; Lu 2011; Ortega, 2003). While some clausal indices were statistically significant, indices on phrasal complexity were better represented in the correlation analysis compared to indices based on clausal complexity. Between the corpora, LOCNESS featured more measures of phrasal complexity that positively correlated with fair score than ICLE-SP. Also, of note was the number of features related to prepositions in LOCNESS. Prepositional phrases allow for more information packing into noun phrases and emerge later in Biber et al.'s (2011) hypothesized sequence of development of grammatical complexity in academic writing. LOCNESS in general features higher scoring essays than ICLE-SP, and therefore likely represents more developed writing. Seven features related to prepositions correlated positively in LOCNESS while there were none in ICLE-SP. Examples 1 and 2 are from the LOCNESS corpus (file code and fair score in parentheses) and feature prepositional phrases (<u>underlined</u>) functioning as adverbials and postnominal modifiers (nouns in bold), respectively. Although prepositional phrases were more prevalent in LOCNESS, they were not completely absent in ICLE-SP. Examples 3 and 4 are from higher scoring essays in ICLE-SP and feature prepositional phrases functioning as postnominal modifiers.

1. *Sartre wrote this play <u>in the middle of a freeze period</u>* (LOCNESS_1, 5.7)
2. *We focus our discussions on money rather than on our assumptions <u>about the value of life</u>* (LOCNESS_196, 4.75)
3. *...collaborate in the construction <u>of a better society</u>...* (ICLE-SP_812, 5.26)
4. *...put at risk the right <u>of a given country to exercise its</u>...* (ICLE-SP_910, 6.56)

With regard to cohesion, only one variable demonstrated a significant correlation with a coefficient of .20 or above in the LOCNESS corpus. As this variable was part of the final regression model for LOCNESS, it will be presented in the next section along with the discussion of the regression models.

This section presented a broad overview of a selection of the statistically significant Pearson correlation coefficients greater than or equal to .20. The full list of these correlations is presented in Appendix 3. In the data, both negative and positive correlations pointed to context specific, less frequent words being more strongly associated with higher fair scores. Furthermore, it was seen that more phrasal than clausal embedding correlated with higher essay scores, and LOCNESS essays had more phrasal embedding features than ICLE-SP. The next section will discuss results of the combination of predictors that accounted for the most variance in predicting fair scores in each corpus.

5.2. Regression analysis

The stepwise regression analyses revealed a combination of six predictors that accounted for 26.7% of the variance in holistic essay scores from ICLE-SP and a combination of 10 predictors that explained 41.4% of the variance in holistic essays scores from LOCNESS. For ease of comparison, those measures that contributed to the models in Tables 3 and 4 are reproduced here side-by-side in Table 5. The minus and plus signs in parentheses next to the variable names indicate negative or positive correlations with fair score.

Table 5 Predictors that contribute to ICLE-SP and LOCNESS stepwise regression models

ICLE-SP	LOCNESS
1. USF_CW (–)	1. Brysbaert_Concreteness_Combined_AW (–)
2. news_av_lemma_construction_freq_log (–)	2. hyper_noun_S1_P1 (+)
3. amod_dobj_deps_NN_struct (+)	3. lsa_2_all_para (+)
4. news_av_lemma_construction_freq_type (–)	4. all_av_lemma_construction_freq_type (–)
5. content_poly (–)	5. amod_pobj_deps_struct (+)
6. SUBTLEXus_Range_FW_Log (–)	6. Brown_Freq_AW (+)
	7. KF_Freq_AW (+)
	8. Phono_N_H (–)
	9. USF_AW (+)
	10. av_dobj_deps_NN (+)

One immediate difference between groups that can be seen in Table 5 is that there is a more restricted set of predictors in ICLE-SP compared to LOCNESS, six and 10, respectively. There is also no direct overlap of predictors between the models, indicating qualitative differences in features that best predict higher holistic writing quality scores between L1 speakers of Spanish and English. While there was no direct overlap, there is common area in terms of the linguistic constructs that the features represent. The most obvious instances of this are the variables that reveal contextual distinctiveness of lexis (e.g., USF_AW), those that cover verb constructions (e.g., news_av_lemma_construction_freq_log), and adjectival modifiers (e.g., amod_pobj_deps_struct). Finally, referring back to Tables 3 and 4, it can be seen that there are more negative correlations between predictor and fair score reflected in the beta figures in ICLE-SP than LOCNESS, five and three, respectively.

Beginning with lexical sophistication measures, fair score in ICLE-SP negatively correlated with words that were contextually broad (USF_CW), highly polysemous (content_poly), and frequent in television and movie subtitles (i.e., common in spoken language). The negative correlation in the first instance indicates that words associated with a broad array of stimuli is predictive of lower scores on essays. This suggests that raters rewarded contextually relevant lexis in the essays. Likewise, common polysemous words such as *think, work,* and *know* will also be common in spoken registers and likely not reflect well in academic writing contexts. In the same vein, fair scores demonstrated a negative correlation with frequent words from the spoken registers of movies and TV series subtitles from the SUBTLEXus corpus.

The lexical sophistication measures that accounted for the most variance in holistic scores in LOCNESS were measures of concreteness, hypernymy, frequency, phonology, and distinctiveness. Examining the nature of the variables, it is plausible that some of these variables overlap or work together in the lexical constructs they represent. For instance, concreteness was negatively correlated with fair score meaning intangible words such as *honesty* or *invention* predicted higher lexical sophistication scores compared to more concrete words with a clear physical referent such as *car* (Kyle & Crossley, 2015). Hypernymy and contextually distinct words were positively correlated variables in the regression model. Higher hypernymy scores suggest more specific and contextually distinct words and thus are likely to be lower frequency words. Finally, the variable of near phonological neighbors (Phono_N_H) indicates words that differ by only one phoneme. Words that have many phonological neighbors are typically shorter, less sophisticated words. Therefore, the negative correlation indicates a lack of sophisticated words in lower scoring essays which in turn implies that higher scoring essays in LOCNESS feature more sophisticated words. Taken together, these measures paint a picture of high scoring essays that feature low-frequency, contextually specific, sophisticated, and often intangible words. For instance, an essay from LOCNESS that received a high fair score of 6.82 features words that fit this profile: *existentialist, guilt, characteristic,* and *duplicity*. Interestingly, there were no features related to multiword units in either regression model. There were, however, numerous predictors on the syntactic plane of language.

Predictors of syntactic complexity in ICLE-SP and LOCNESS included measures of main verb lemmas that frequently combine with VACs in COCA (Davies, 2008-)

and were negatively correlated with fair score (e.g., all_av_lemma_construction_freq_type). Common main verb lemmas in this measure include *be, say, have, make, go, get,* and *take*. It is thus not surprising that these predictors are again negatively correlated with fair score as raters of academic essays are likely to reward usage of less frequent, more descriptive verbs such as *expounded* rather than a more common verb such as *said*. This finding is in line with Kyle and Crossley's (2017) research in which negative correlations of similar variables were found in learner writing.

The only positive predictor related to syntactic complexity in the ICLE-SP model was the variable for adjectives modifying a noun phrase, excluding pronouns (amod_dobj_deps_NN_struct). Likewise, the LOCNESS corpus has two predictors related to noun phrase complexity indicating that, in general, authors from the LOCNESS corpus use a wider variety of phrasal embedding in noun phrases than those authors represented in the ICLE-SP corpus. It is also of note that the effect of these predictors is quite pronounced relative to the other positive predictors in the model via comparisons of the unstandardized beta coefficients. The variable amod_pobj_deps_struct, representing adjectival modifiers, had the largest effect among predictors featuring a positive correlation with fair score and an unstandardized beta value of 2.39.

With regard to measures of cohesion, surprisingly, no variables in ICLE-SP had a significant positive correlation greater than or equal to .20 with fair score. In LOCNESS, only the variable (lsa_all_para) was a significant predictor. This variable represents intra-text semantic relations attained via latent semantic analysis and provides support to Crossley et al.'s (2016) finding that measures of global cohesion positively predict overall cohesion scores. One reason for the minimal numbers of cohesion indices found here may be less about a paucity of cohesive features, and more about the grouping of the variables. For example, TTR was grouped with the lexical measures as it captures lexical diversity, but many scholars group TTR with cohesion as it can be used to quantify repetition of words and thus capture cohesive properties. Likewise, measures of hypernymy could arguably be included with cohesive measures as hypernyms and hyponyms can result in complex lexical chains that traverses a text while displaying linguistic diversity. In essence, many of these variables exist along a continuum in terms of the language proficiency constructs they are deemed to represent, but they are forced into a single category for analysis.

6. Practical implications

Findings from this study can be used to inform teaching in terms of features that tend to result in higher essay scores. Students should be aware that holistically scored essays do indeed reflect an array of linguistic constructs, and that essay scores appear to increase when more of these constructs are represented. Evidence presented here supports the inclusion of more obvious features in essays such as using more sophisticated and contextually specific vocabulary appropriate for written registers. Less obvious features that could be beneficial to learners are related to syntactic and phrasal complexity via embedding adjectives and prepositional phrases in noun phrases. This finding is congruent with research on academic writing (e.g., Biber et al., 2011; Parkinson & Musgrave, 2014), but may be lesser known by practitioners in EFL and ESL contexts. For instance, using adjectives to describe nouns immediately adds a layer of phrasal complexity as well as creating more descriptive and engaging writing. Example 5 illustrates a higher scoring essay from ICLE-SP that features prepositional phrases as post-nominal modifiers (underlined) and attributive adjectives (adjectives are double underlined). This example also clearly utilizes sophisticated and contextually specific vocabulary. Just as nouns take modifying prepositional phrases for embedding, so can adjectives as in Example 6.

5. *... the former brutality of capitalism has disappeared or, at least, has been reduced to marginal groups...* (ICLE-SP_805, 5.24)
6. *... the British electorate, which still seems wary about the idea of a Britain fully incorporated into Europe...* (LOCNESS_15, 6.14)

Finally, students could be taught the value of complex repetition via synonymy and hypernymic relations to demonstrate a broader and deeper lexicon and for building more cohesive essays. Language instructors could encourage the use of thesauri to find synonyms and internet-based resources such as WordNet (2010) to find hyper- and hyponyms to increase lexical diversity and complex repetition in writing. Taking this array of language proficiency constructs into account when writing essays, in addition to focusing on compelling content, should be of benefit to learners aiming to increase their writing quality scores.

7. Conclusion

The present study used correlation analysis and a stepwise regression to investigate the extent to which a broad set of writing features covering lexical complexity, syntactic complexity, and cohesion explained the variance in holistic scores of English essays authored by L1 speakers of Spanish and English. Texts from the ICLE-SP version 2 and LOCNESS corpora were holistically rated by a group of trained raters and subjected to a Many Facet Rasch Measurement to arrive at fair scores for each essay. The essays were then analyzed using custom scripts and the NLP software packages TAALES, TAASSC, and TAACO to generate indices of lexical complexity, syntactic complexity, and cohesion, respectively. A wide variety of measures significantly correlated with fair scores for both groups of speakers. For instance, the correlation analysis illustrated that phrasal complexity correlated with high scores in both groups, but the LOCNESS group showed more correlations with phrasal complexity grounded in prepositions, a finding that supports the work of Biber and colleagues. The regression models for both groups suggest that the use of less frequent, contextually specific words in writing marked by phrasal complexity will better predict higher essay scores. However, exact predictors and the nature of those predictors differed. In the case of ICLE-SP, predictors by and large revealed negative correlation coefficients, suggesting features and trends to avoid in order to attain higher writing quality scores. Texts in LOCNESS, on the other hand, featured more positive correlation coefficients in its explanatory variables such as embedding in noun phrases via adjective and prepositional phrases, and global cohesion.

Of course, there are limitations to this study. NLP tools are powerful and allow for the analysis of large data sets that can potentially pick up on trends that might go unnoticed by the human eye. However, even advanced tools are hampered by the state of the art. For example, it may be the case that there are cohesive devices functioning throughout the essays that affect scores, but the way a tool operationalizes cohesion may not align with human sensitivity to cohesion. More importantly, the present analysis forgoes analysis of the content of the essays, focusing only on surface level features. Raters are most certainly affected by content choices writers make, but the automated methods used in this analysis do not directly measure or judge content. Therefore, it is necessary to bear in mind that the indices reported here only account for a percentage of the variance in the essay scores. Lastly, and probably of most consequence, topic was not accounted for in the regression model. Undoubtedly, certain topics prompted different lexical choices and syntactic patterns. While this scenario is likely, it may not be cause for too much concern as there was overlap of five topics between the corpora, and all topics will result in vocabulary specific to the topic. Ultimately, the findings in this paper reveal trends that are likely of relevance even when factoring in the effect of topic.

Despite these limitations, the analysis conducted here contributes to the literature in at least two ways. First, it provides support to previous studies on learner language aimed at lexical complexity, syntactic complexity, and cohesion by using a data set that had not previously been scored for writing quality. Studies on different data sets help to validate previous findings and allow for more confidence in the inferences gleaned from a larger body of research. Secondly, by using measures from different linguistic proficiency constructs, this study helped to fill a gap in broad measures of writing indices that contribute to holistic writing quality scores. It is hoped that the findings presented here can be of use to students and teachers in writing contexts.

References

Berger, C. M., Crossley, S., & Kyle, K. (2019). Using native-speaker psycholinguistic norms to predict lexical proficiency and development in second language production. *Applied Linguistics, 40*(1), 22-42. https://doi.org/10.1093/applin/amx005

Biber, D. & Conrad, S. (2009). *Register, genre, and style.* Cambridge University Press.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405. https://doi.org/10.1093/applin/25.3.371

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: complexity, elaboration, explicitness. *Journal of English for Academic Purposes, 9*(1), 2-20.

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly, 45*(1), 5-35. https://doi.org/10.5054/tq.2011.244483

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2013). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904-911. https://doi.org/10.3758/s13428-013-0403-5

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26,* 42-65. https://doi.org/10.1016/j.jslw.2014.09.005

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397-423. https://doi.org/10.1016/j.esp.2003.12.001

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(1), 213-238. https://doi.org/10.2307/3587951

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods 48*(4), 1227-1237. https://doi.org/10.3758/s13428-015-0651-7

Crossley, S. A., Salsbury, T., & McNamara, D. (2010). The development of polysemy and frequency use in English second language speakers. *Language Learning, 60*(3), 573-605. https://doi.org/10.1111/j.1467-9922.2010.00568.x

Davies, M. (2008-) *The corpus of contemporary American English (COCA)*. Available online at https://www.english-corpora.org/coca/.

Eckes, T. (2015). *Introduction to many-facet Rasch measurement: Analyzing and evaluating rater-mediated assessments* (2nd ed.). Peter Lang.

Ellis, N. C., & Ferreira-Junior, F. (2009). Construction learning as a function of frequency, frequency distribution, and function. *Modern Language Journal, 93*(3), 370-385.

González, M. C. (2017). The contribution of lexical diversity to college level writing. *TESOL Journal, 8*(4), 899-919. https://doi.org/10.1002/tesj.342

Granger, S. (Ed.) (1998). *Learner English on computer.* Longman.

[LOCNESS]. http://www.learnercorpusassociation.org/resources/tools/locness-corpus/

Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (Eds.). (2009). *ICLE: International corpus of learner English. Louvain-la-Neuve*. Presses Universitaires de Louvain.

Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics, 13,* 403-437. https://doi.org/10.1075/ijcl.13.4.02gri

Halliday, M. A. K., & Hasan. R. (1976). *Cohesion in English.* Longman.

Hoey, M. (1991). *Patterns of lexis in text*. Oxford University Press.

Kim, M., Crossley, S. A., & Kyle, K. (2018). Lexical sophistication as a multidimensional phenomenon: Relations to second language lexical proficiency, development, and writing quality. *Modern Language Journal, 102*(1), 120-141. https://doi.org/10.1111/modl.12447

Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication* (Doctoral Dissertation). http://scholarworks.gsu.edu/alesl_diss/35

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: indices, tools, findings, and application. *TESOL Quarterly 49*, 757-786. https://doi.org/10.1002/tesq.194

Kyle, K., & Crossley, S. A. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. *Language Testing, 34*(4), 513-535. https://doi.org/10.1177/0265532217712554

Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *Modern Language Journal, 102*(2), 333-349. https://doi.org/10.1111/modl.12468

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*, 307-322. https://doi.org/10.1093/applin/16.3.307

Linacre, J. M. (2018). *Facets computer program for many-facet Rasch measurement* (version 3.80.4). Winsteps.com

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writer's language development. *TESOL Quarterly, 45*, 36-61. https://doi.org/10.5054/tq.2011.240859

Meara, P. (2005). Designing vocabulary tests for English, Spanish and other languages. In C. Butler, S. Christopher, M. Á. Gómez González, & S. M. Doval-Suárez (Eds.), *The dynamics of language use* (pp. 271-285). John Benjamins.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24*(4), 492-518. https://doi.org/10.1093/applin/24.4.492

Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes, 14,* 48-59. https://doi.org/10.1016/j.jeap.2013.12.001

Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.

Schanding, B. & Pae, H. K. (2018). Shell noun use in argumentative essays by native speakers of Japanese, Turkish, and English: Frequency and rate of noun pattern attraction. *International Journal of Learner Corpus Research, 4*(1), 54-81. https://doi.org/10.1075/ijlcr.16014.sch

Simpson-Vlach, R. & Ellis, N. C. (2010). An academic formula list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487-512. https://doi.org/10.1093/applin/amp058

Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication, 33*, 149-183. https://doi.org/10.1177/0741088316631527

WordNet. (2010). Princeton University „About WordNet." https://wordnet.princeton.edu/.

Yoon, H. J. (2018). The development of ESL writing quality and lexical proficiency: Suggestions for assessing writing achievement. *Language Assessment Quarterly, 15*(4), 387-405. https://doi.org/10.1080/15434303.2018.1536756

APPENDIX 1

Topics in ICLE-SP and LOCNESS

	ICLE-SP	LOCNESS
1	General Social Issues	
2	Crime and Punishment	
3	Environmental Issues	
4	Reflections on Literature	
5	Feminism	
6	University	French culture
7	Technology Stifles the Imagination	A single Europe: a loss of sovereignty in Britain
8	Television is the Opiate of the Masses	Education, language, and policy
9	Types of Theatre	Money is the root of all evil
10	Military Service	Media
11		Sports, leisure, and hobbies
12		Family issues
13		Health

APPENDIX 2

Holistic writing rubric

7: Uses cohesion in such a way that it attracts no negative attention. Skillfully manages paragraphing. Uses a wide range of vocabulary with sophisticated and natural control. Uses a wide range of grammatical structures accurately and with flexibility.

6: Sequences information and ideas logically. Manages aspects of cohesion well and uses paragraphing sufficiently and appropriately. Uses a wide range of vocabulary, including uncommon words, skillfully and flexibly but there may be occasional inaccuracies. Uses a wide range of grammatical structures, usually error free.

5: Logically organizes information and there is clear progression. Uses a range of cohesive devices appropriately though there may be some under-/over-use. Uses a sufficient range of vocabulary, uses less common lexical items, may have occasional errors in word choice or spelling. Uses a variety of complex structures, produces frequent error-free sentences, has good control of grammar and punctuation but may have a few errors.

4: Arranges ideas coherently and there is clear overall progression. Uses cohesive devices but cohesion between sentences may be faulty. May not always use referencing clearly or appropriately. Adequate range of vocabulary, attempts to use less common words but with some inaccuracy. Uses a mix of simple and complex sentence forms, makes some errors in grammar and punctuation but this rarely reduces meaning.

3: Presents information with some organization but may lack overall progression. Inadequate, inaccurate, or overuse of cohesive devices. May be repetitive due to lack of referencing and substitution and paragraphing may be inadequate. Vocabulary is limited with noticeable errors in spelling and word formation that cause difficulty for the reader. Uses only a limited range of grammatical structures. Attempts complex sentences but they tend to be inaccurate. Frequent grammatical and punctuation errors can cause some difficulty for the reader.

2. Presents information and ideas but these are not arranged coherently and there is no clear progression in the response. Uses some basic cohesive devices these may be inaccurate and paragraphing may be absent or confusing. Uses only basic vocabulary which may be repetitive or inappropriate for the topic. Limited range of structures with only rare use of subordinate clauses. Errors predominate structures and punctuation is faulty.

1. Does not organize ideas logically. May use a very limited range of cohesive devices and those used may not indicate a logical relationship between ideas. Uses only a very limited range of words and expressions with very limited control of word formation and/or spelling. Errors may distort message. Attempts sentences but errors in grammar and punctuation distort meaning.

APPENDIX 3

Summary of indices with a correlation coefficient of 0.200 or greater with essay fair score in ICLE-SP and LOCNESS

ICLE-SP

Type of measure	Variable Name	Correlation Coefficient
Lexical complexity	Essay_TTR	0.242
	lexical_density_types	0.202
	KF_Freq_AW_Log	-0.228
	KF_Ncats_AW	-0.276
	KF_Nsamp_AW	-0.236
	TL_Freq_AW_Log	-0.271
	Brown_Freq_AW_Log	-0.260
	TL_Freq_CW	-0.228
	MRC_Familiarity_AW	-0.327
	MRC_Meaningfulness_AW	-0.274
	MRC_Familiarity_CW	-0.315
	MRC_Meaningfulness_FW	-0.239
	Kuperman_AoA_AW	0.341
	Brysbaert_Concreteness_Combined_AW	-0.226
	Brysbaert_Concreteness_Combined_FW	-0.210
	SUBTLEXus_Range_AW	-0.326
	SUBTLEXus_Freq_CW	-0.277
	SUBTLEXus_Range_FW_Log	-0.235
	BNC_Spoken_Bigram_Normed_Freq_Log	-0.250
	McD_CD	0.217
	USF_CW	-0.291
	McD_CD_CW	0.292
	All_AWL_Normed	0.268
	PLD	0.311
	PLDF	-0.312
	Ortho_N_CW	-0.341
	PLDF_CW	-0.316
	LD_Mean_RT_Zscore	0.299
	LD_Mean_RT_SD	0.223
	LD_Mean_Accuracy	-0.212
	WN_Mean_RT	0.354
	LD_Mean_RT_SD_CW	0.200
	LD_Mean_Accuracy_CW	-0.223
	aoe_inverse_linear_regression_slope	0.206
	aoe_inflection_point_polynomial	0.255
	content_poly	-0.243
	hyper_verb_S1_P1	0.218
Syntactic complexity	MLC	0.233
	CN_C	0.207
	av_nominal_deps	0.217
	av_nominal_deps_NN	0.202
	av_dobj_deps_NN	0.217
	amod_all_nominal_deps_struct	0.210

	amod_dobj_deps_NN_struct	0.229
	news_av_lemma_construction_freq_log	-0.239
	news_av_lemma_construction_freq_type	-0.202
	fic_av_lemma_freq_log	-0.283

LOCNESS

Type of measure	Variable Name	Correlation Coefficient
Lexical complexity	lemma_ttr	-0.209
	lexical_density_types	0.283
	KF_Freq_AW	0.208
	Brown_Freq_AW	0.274
	KF_Ncats_CW	-0.238
	TL_Freq_CW	-0.205
	MRC_Concreteness_AW	-0.276
	MRC_Meaningfulness_AW	-0.295
	MRC_Familiarity_CW	-0.266
	Kuperman_AoA_CW	0.340
	Brysbaert_Concreteness_Combined_AW	-0.285
	Brysbaert_Concreteness_Combined_CW	-0.270
	SUBTLEXus_Freq_CW	-0.215
	SUBTLEXus_Range_CW_Log	-0.288
	USF_AW	-0.200
	McD_CD	0.259
	USF_CW	-0.215
	McD_CD_CW	0.272
	eat_tokens_FW	0.202
	Ortho_N	-0.258
	Phono_N_H	-0.264
	Phono_N_H_CW	-0.265
	OLD_CW	0.278
	LD_Mean_RT_SD	0.222
	LD_Mean_Accuracy	-0.263
	WN_SD	0.258
	WN_Mean_Accuracy_CW	-0.314
	hyper_noun_S1_P1	0.215
	hyper_verb_S1_P1	0.243
	hyper_verb_noun_s1_p1	0.258
Syntactic complexity	av_nominal_deps	0.273
	av_pobj_deps	0.246
	av_nominal_deps_NN	0.250
	av_dobj_deps_NN	0.269
	av_pobj_deps_NN	0.207
	nominal_deps_stdev	0.233
	nsubj_NN_stdev	0.212
	nsubj_pass_NN_stdev	0.208
	pobj_NN_stdev	0.217
	amod_all_nominal_deps_struct	0.240
	prep_all_nominal_deps_NN_struct	0.271
	amod_dobj_deps_struct	0.205
	prep_dobj_deps_NN_struct	0.203

	amod_pobj_deps_struct	0.227
	prep_pobj_deps_struct	0.253
	fic_av_lemma_freq_log	-0.218
	all_av_lemma_construction_freq_type	-0.233
	acad_av_faith_verb_cue_stdev	0.239
	fic_av_faith_verb_cue_stdev	0.211
	MLS	0.244
	MLC	0.230
	CN_T	0.301
	CN_C	0.278
Cohesion	lsa_2_all_para	0.234