The Asian Journal of Applied Linguistics
Vol. 7 No. 1, 2020, pp. 1-17

AJAL

Emotion and syntactic complexity in L2 writing: A corpus-based study on Chinese college-level students' English writing

Yizhou Wang
The University of Melbourne, Australia

This study quantitatively investigates the role of emotion in L2 writing based on corpus analyses using automatized tools. Over two thousand six hundred essays written by Chinese college-level EFL writers were selected from the TECCL corpus (Xue, 2015) and analyzed using both emotionality and syntactic complexity information extracted from the written texts. Regression analysis revealed that while Chinese EFL learners tended to write positively overall, positive writing prompts generally lead to higher emotional scores in their written responses (r = 0.351, p < 0.001). Two clause-level complexity indices have shown an emotional effect, and the highest complexity scores were found when the textual emotion was neutral, while both positive and negative emotions during writing were associated with a lower score in the indices of mean length of clause (MLC) (p < 0.01), and coordinate phrases per clause (CP/C) (p < 0.01). Correlation and dimensionality analyses raise questions about the original grouping method of the fourteen complexity indices proposed by Lu (2010), as the indices from each group did not yield reliable measurements. Overall, the results suggest that emotion may play an important role in syntactic complexity in L2 writing, which should be taken into consideration in language teaching and assessment.

Keywords*:* EFL; L2 writing; emotion; syntactic complexity; corpus; Chinese learners

Introduction

Emotion plays a unique and irreplaceable role in language learning and development (Dewaele & Pavlenko, 2002; Gabryś-Barker, 2009; Schumann, 1994; Swain, 2013), as well as in the general domain of cognitive processing (Bach & Dayan, 2017; Dolan, 2002; LeDoux, 1989). Neurologically, emotion is highly related to the processing mechanisms of the amygdala (Davis & Whalen, 2001), an almond-shape set of neurons that are located in the medial temporal lobe. From the perspective of evolution, human emotion is generally considered as an important algorithm that assists organisms in making optimal decisions in different survival scenarios (Bach & Dayan, 2017). For example, activation of the amygdala is detected when a human subject is perceiving a potential threat and experiencing a negative emotion (Breiter et al., 1996). Specifically, for foreign or second language (L2) learning, emotion is considered part of the general cognitive framework of high-level constructs for second language acquisition (SLA), e.g., the affective filter hypothesis (Du, 2009; Krashen, 1981; Schumann, 1994). Empirical evidence has shown that learners' emotional status can be associated with their L2 performance. For instance, Pishghadam (2009) reported that successful foreign-language writers have better skills in stress management and mood adaptability, compared to less successful learners.

Research has also shown that emotional factors in L2 writing can influence learners' cognitive distribution among different linguistic aspects, e.g., lexical and morphosyntactic *vs.* pragmatic planning (Clachar, 1999), as well as the end-point text quality, e.g., syntactic and content structure (Kean, Glynn, & Britton, 1987). For instance,

ISSN 2308-6262
http://caes.hku.hk/ajal

Clachar (1999) analyzed the language planning process of L2 learners during emotional and non-emotional composition using a thinking-aloud protocol. It was found that emotional prompts were associated with a higher proportion of planning activities dedicated to pragmatic factors in non-emotional (18.62%) than emotional composition (11.25%). The activity proportion of textual planning was also higher in the non-emotional (39.14%) than in the emotional task (32.72%). Finally, L2 learners spent more time in processing lexico-morpho-syntactic information when provided the emotional prompt (55.83%), compared to the non-emotional prompt (42.24%). What is unknown is whether the emotionality of the learners' written response can influence the quality of linguistic components in their writing. In an earlier regression study (Kean et al., 1987), student writers' anxiety level was found to be negatively correlated with the quality of their writings ($r = -0.26$, $p < 0.05$), suggesting that the emotional status can influence student writers' composition performance. However, since the anxiety level was measured apart from the writing task, it was unclear whether the anxiety level was stable during the writing process. While there is limited L2 writing research on the role of emotion conducted in the Asian context, the present study focuses on the role of emotion in L2 English writing performance by Chinese college-level students who learn English as a foreign language (EFL). In particular, the present study quantitatively analyses the effect of emotion on the syntactic complexity in EFL learner's writing performance using an automated analyzer L2SCA (Lu, 2010) based on a large local corpus of 2,620 texts, which are selected from a pre-collected corpus (Xue, 2015).

Emotional states can be assessed using different instruments, including qualitative clinical discourse analysis (e.g.Vaughn & Leff, 1976), standardized scales (e.g. Gratz & Roemer, 2004; Weinberg & Klonsky, 2009), behavioural observation (e.g. Bradley & Lang, 2000; Mauss, Levenson, McCarter, Wilhelm, & Gross, 2005), and neurophysiological correlates (Kissler, Herbert, Winkler, & Junghofer, 2009; Schupp, Junghöfer, Weike, & Hamm, 2003). In addition, sentiment analysis, as a text-based emotion detection tool, has been widely used to extract emotional information in the area of computer and language science (Balahur, Mihalcea, & Montoyo, 2014; Kaur & R. Saini, 2014; Liu, 2015), and is recently suggested as an complementary instrument in clinical intervention for depression treatment (Provoost, Ruwaard, van Breda, Riper, & Bosse, 2019). As an automatized emotion detector, sentiment analysis can identify and quantify the emotional information in the text in terms of sentiment polarity (positive, neutral, or negative). Words that have positive valences (e.g., *happy*) will add up to the total score of the positivity score, while negative words (e.g., *sad*) will increase the negativity score. A resultative value of sentiment score can be derived from subtracting the negativity from the positive score. A positive (+) value of the sentiment score indicates that the global text emotion is positive, while a negative (-) score means that the overall emotional status is negative. Sentiment analysis can be a useful tool for detecting the emotional states of the L2 writers during composition, especially when treating corpus-based text data.

In L2 writing research, syntactic complexity is generally considered as an essential aspect of writing quality, as well as a robust measurement of L2 proficiency (Ortega, 2003; Taguchi, Crawford, & Wetzel, 2013; Vyatkina, 2013). Previous studies have also found that syntactic complexity in L2 writing can be influenced by many factors, e.g., time pressure (Ellis & Yuan, 2004), prompt genre (Way, Joiner, & Seaman, 2000) and topic choice (Yang, Lu, & Weigle, 2015). To measure syntactic complexity, researchers generally define it as a multi-dimensional construct and adopt a variety of indices as effective measures for different linguistic dimensions (Bulté & Housen, 2012; Norris & Ortega, 2009).

In particular, a series of studies using automatized rating algorithms (Lu, 2010, 2011, 2017; Lu & Ai, 2015) have proposed a consistent framework of fourteen indices for measuring syntactic complexity in L2 writing (Table 1). The fourteen indices are designed to measure syntactic complexity in five linguistic dimensions: (1) length of production unit, (2) amount of subordination, (3) amount of coordination, (4) degree of phrasal sophistication, and (5) overall sentence complexity. With the automatized analyzer, L2SCA (Lu, 2010), the fourteen indices can be efficiently computed on a large text corpus. While Lu's taxonomy represents a two-level structure (five dimensions and fourteen measures), Yang et al. (2015) have recently proposed that syntactic complexity should be modelled as a hierarchical structure. If Lu's model provides an effective taxonomy of the syntactic dimensions, it is expected that the indices within each syntactic dimension should show consistent patterns quantitatively, e.g., these indices should have good reliability measures and correlational properties. While theoretical discrepancies exist, there is limited evidence whether L2SCA indices can reliably test the linguistic features in the corresponding dimension when tested on a large dataset. Apart from applying the automatic analysis with the L2SCA tool, this study also sets out to critically evaluate the performance of the analyzer in terms of its measurement reliability in different syntactic dimensions.

The present study aims to answer three research questions: (1) Given the evidence that emotional writing prompts may influence L2 writers' cognitive distribution during composition (Clachar, 1999), does the emotionality of the writing topics affect the emotionality of the writers' productions? (2) Since the writing topic was previously found to be an important factor that influences L2 writer's syntactic complexity (Yang et al., 2015), does the emotionality of the writing prompts influence syntactic complexity in L2 writing overall? (3) Do the fourteen indices proposed by Lu (2010) reliably measure text complexity in the five linguistic dimensions? The method of the present study adopts corpus analyses, and the scope of the research findings are limited to the quantitative results based on the dataset.

Table 1. Lu's (2010) fourteen indices for L2 syntactic complexity

Dimension	Measure/Index	Code
Length of production unit	Mean length of clause	MLC
	Mean length of sentence	MLS
	Mean length of T-unit	MLT
Amount of subordination	Clauses per T-unit	C/T
	Complex T-unit per T-unit	CT/T
	Dependent clauses per clause	DC/C
	Dependent clauses per T-unit	DC/T
Amount of coordination	Coordinate phrases per clause	CP/C
	Coordinate phrases per T-unit	CP/T
	Verb phrases per T-unit	T/S
Degree of phrasal sophistication	Complex nominals per clause	CN/C
	Complex nominals per T-unit	CN/T
	Verb phrases per T-unit	VP/T
Overall sentence complexity	Clauses per sentence	C/S

Method

Corpus data

The present study uses 2,620 essays written by Chinese college-level EFL learners from the Ten-thousand English Compositions of Chinese Learners (TECCL) corpus (Xue, 2015). Originally, the TECCL corpus contains 1.8 million words in around 10,000 texts, which are written by Chinese EFL learners at different educational levels and were submitted from 2010 to 2015. Over 1,000 different essay prompts were collected in the corpus, which covers a wide range of topics, and the text materials were collected from 32 provinces within China. For the purpose of the present study, the selection covers 2,620 written texts by college-level writers from six regions: Beijing (n = 619), Guangdong (n = 435), Hebei (n = 238), Jiangsu (n = 706), Shanghai (n = 247), and Zhejiang (n = 375). Most of the writing samples were short essays used as academic English training tasks in colleges and universities with an average length of 200 words (SD = 125 words). While the exact genres of the text were not already classified in the corpus, the stylistic features seem to resemble analytical and argumentative writings. For instance, many writing prompts are contextualized in contemporary social issues, e.g., excessive packaging, food safety and low-carbon lifestyle. The TECCL corpus also marks Chinese elite universities, i.e., the so-called 985-211 project universities (Fang, 2012; Wu, 2015), and the selected local corpus has 205 (7.8%) essays from those universities, while the majority (n = 2415, 92.2%) are from non-elite college-level institutions. Few studies have analyzed Chinese EFL writers' L2 writing performance on a comparable scale to the present study.

Sentiment analysis and L2 syntactic complexity analysis

Emotional information from the texts, i.e., both prompts and essays, were extracted from a series of sentiment analyses conducted in R (R Core Team, 2019) using the *SentimentAnalysis* package (https://github.com/sfeuerriegel/SentimentAnalysis). In the current analysis, the sentiment valency for each text is computed using the *Harvard-IV Dictionary*, which is a psychological dictionary published by Harvard University (http://www.wjh.harvard.edu/~inquirer/) and was used in the *General Inquirer* software (Stone, Bales, Namenwirth, & Ogilvie, 1962). When coding sentiment (emotion) scores, the positive values are encoded as positive, the negative values are encoded as negative, and the zero scores are coded as neutral. Within the local corpus, 1,067 prompts are judged as positive, 517 topics are negative, and 1,036 topic entries are neutral. Typical negative prompts include *Water Crisis*, *City Problems* and *Inequality*. Typical positive prompts are generally associated with positive connotations , e.g., *Enjoying learning English online*, *How to make a good impression?* and *To be an optimistic person*. Neutral prompts have zero sentiment scores, e.g., *Compete or Cooperate*, *Online Shopping* and *My Campus Life*. For the writing responses, the majority (n = 2,421) of the texts are judged as positive, 148 texts are judged as negative, while 51 essays are encoded as neutral. The fourteen L2SCA syntactic complexity indices (Lu, 2010) were automatically computed using the analyzing software TAASSC (Kyle, 2016), which offers a graphic interface for importing and exporting materials.

Data analysis

To address research question one, the correlation between the emotion scores of prompts and written texts were analyzed using a Pearson regression. The second research question was approached using a series of mixed-effects models (Magezi, 2015). The mixed-effects models take syntactic complexity indices as the dependent variable, while the sentiment types of both the prompt and the written text are taken as fixed predictors. Since the essays in the analysis come from only six regions in China, the categorical variable *region* is taken as a random factor. Similarly, the year of submission is also taken as a random factor. Any significant result will suggest an emotional effect on syntactic complexity. The third research question was addressed using a pairwise correlation and then a series of reliability tests. An additional principal components analysis (PCA) was used to check the inner dimensionality of the fourteen indices based on their quantitative similarities.

Results

Emotion in prompts and written texts

Sentiment scores were computed for both the task prompts and written responses in the dataset, where positive (+) values indicate positive emotionality, and negative (-) values indicate negative sentiment in the text. For the whole dataset (N = 2620), the mean sentiment score of the prompts equalled 0.129, while the written texts had a mean sentiment score of 0.094. Both mean values were above zero (Figure 1a). When checked by a one-sample t-test, the mean sentiment score for the prompts was significantly different from the zero baseline, $t(2619) = 13.524$, $p < 0.001$, Cohen's $d = 0.246$. The mean score for the response texts was also significantly different from the zero baseline, $t(2619) = 72.530$, $p < 0.001$, Cohen's $d = 1.417$.

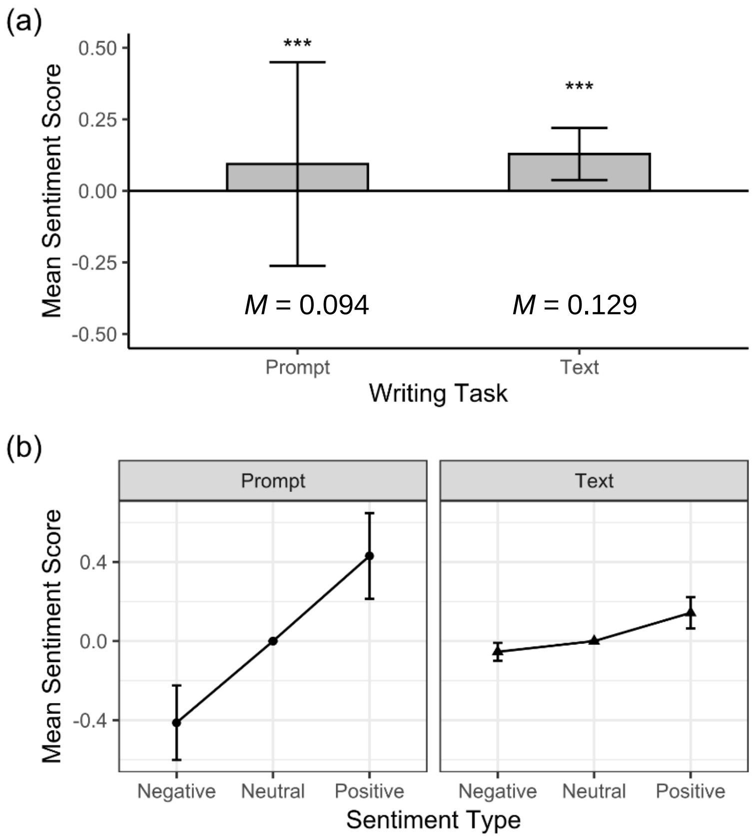

Figure 1a. Mean sentiment scores of the prompts and the written texts in the dataset
Figure 1b. Mean scores for different sentiment types in the prompts and written texts
Note: Bars represent standard deviations

When collapsed into tripartite categories, the sentiment scores for each text type for the writing prompt and response text are summarized in Figure 1b. The sentiment score for the prompts seems to be more polarized. For the negative prompts, M = -0.413, SD = 0.188, range = -1 ~ -0.111; For the positive prompts, M = 0.431, SD = 0.217, range = 0.077 ~ 1. The writing response texts seem to have a less polarized score structure. For negative texts, M = -0.054, SD = 0.045, range = -0.250 ~ -0.003; For positive texts, M = 0.143, SD = 0.079, range = 0.003 ~ 0.491. The sentiment types of the prompt and text were used as predictors when analysing the syntactic indices. The overall interaction between the prompt sentiment and the text emotion was checked by a Pearson's correlation, r = 0.351, p < 0.001 (Figure 2). A positive correlation was found. This suggests that the textual emotion can be effectively predicted by the emotionality of the writing prompts, and EFL college writers tend to write more positively if the writing prompt has positive implications.

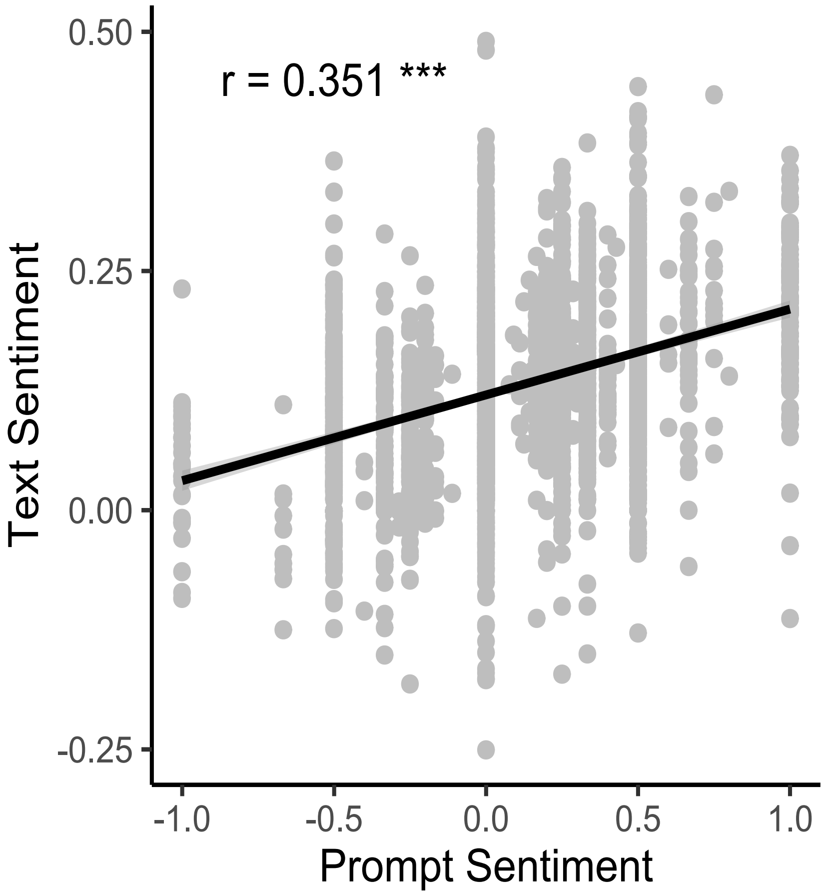

Figure 2. Scatter plot of the sentiment score of the written texts and the prompts

Sentiment type and syntactic complexity

All fourteen indices of syntactic complexity were analyzed using linear mixed models taking tripartite (sentiment score, > 0, positive, = 0, neutral, and < 0, negative) factors *prompt type* and *text type* as fixed effects, while *submission year* and *region* are taken as random factors. The detailed modelling results for all the indices are summarized in the appendix. For the indices that measure the length of production unit (MLC, MLS, and MLT), only MLC is found significant for both prompt type, $F(2, 2611) = 3.5008$, $p = 0.0302$, and text type, $F(2, 2611) = 5.1642$, $p = 0.0057$, but the interaction is not significant, $F(4, 2611) = 1.0128$, $p = 0.3994$. For MLC, unaffected writing responses (text type = neutral) is significantly different from the baseline, $\beta = 2.4425$, $SE = 0.6701$, $p < 0.001$, which indicates that Chinese EFL writers tend to have a higher score for mean length of clauses when they write unemotionally. Based on the current corpus, MLC has the highest mean score when the texts are neutral, $M = 10.804$, $SD = 8.099$. The mean score is somewhat lower when the texts are judged as negative, $M = 9.044$, $SD = 2.077$, or positive, $M = 9.542$, $SD = 2.718$ (see Figure 3a). This pattern is also confirmed by a one-way ANOVA test, $F(2, 2617) = 7.075$, $p < 0.001$. Post hoc Tukey tests reveal that the difference is significant between the neutral and positive texts, $p = 0.006$, and between the neutral and negative texts, $p < 0.001$, but not between positive and negative texts, $p = 0.104$. Similarly, the ANOVA test concerning prompt type alone also yielded significant differences between the three levels, $F(2, 2617) = 4.705$, $p = 0.009$. Post hoc tests revealed that the difference is significant between neutral and negative prompts, $p = 0.013$, and close to significant between neutral and positive prompts, $p = 0.071$.

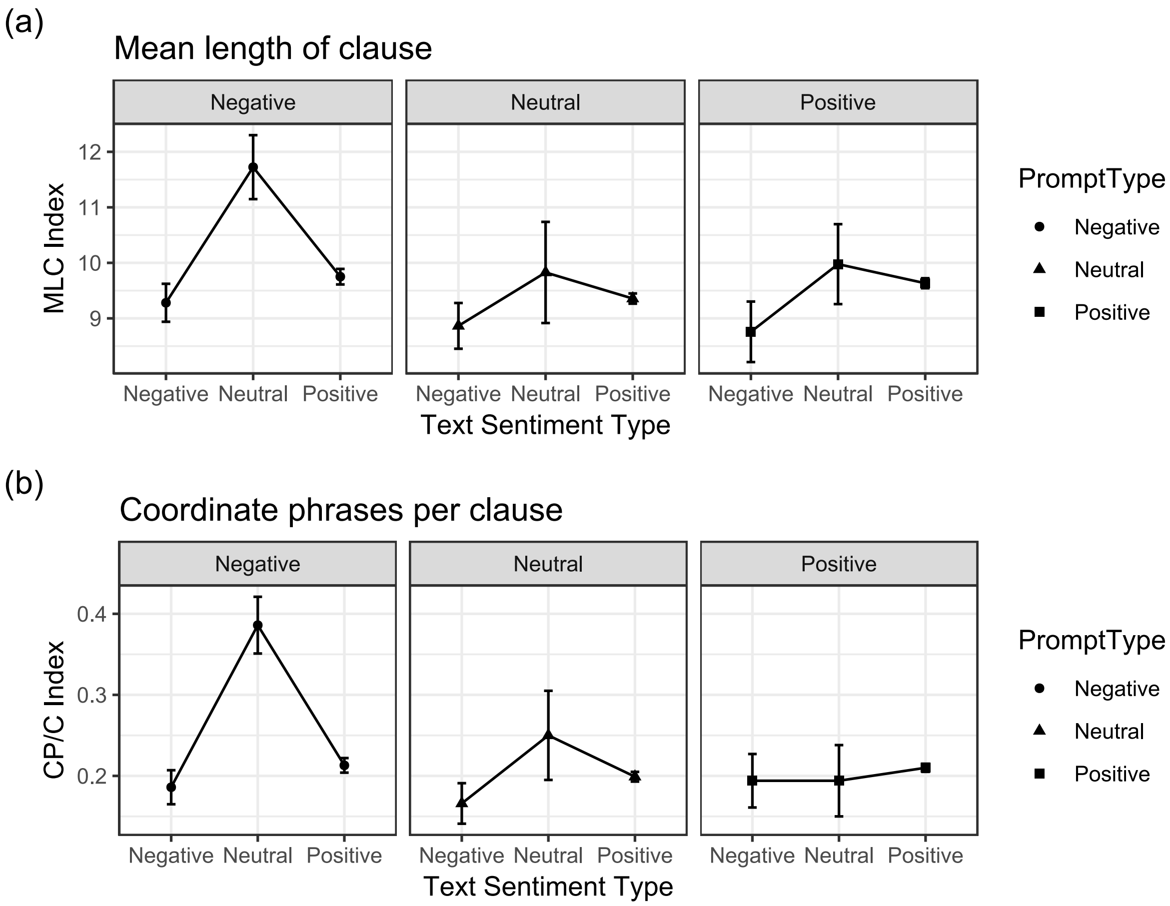

Figure 3a. The score of the MLC index at different prompt type and text type levels
Figure 3b. The score of the CP/C index at different prompt type and text type levels
Note: Bars represent standard errors of the mean

The index of coordinate phrases per clause (CP/C) also reached a significant level for prompt type, $F(2, 2601.9) = 4.7739$, $p = 0.0085$, and text type, $F(2, 2580.0) = 4.8865$, $p = 0.0076$, as well as the interaction, $F(4, 2607.2) = 2.9513$, $p = 0.0190$. In the linear mixed model, the coefficient of neutral text type is a significant predictor, $\beta = 0.2010$, $SE = 0.0407$, $p < 0.001$, as well as the interaction between positive prompts and neutral texts, $\beta = -0.2005$, $SE = 0.0684$, $p = 0.0034$. Pairwise contrastive tests using Tukey adjustments revealed that the estimated marginal mean (EMM) between negative and positive prompts is significant, $p = 0.0177$, but the interaction difference is not significant between negative and neutral prompts, $p = 0.0599$, or between neutral and positive prompts, $p = 0.9647$. As for text types, the marginal mean value is not significant between the three levels, but close to significant between negative and neutral texts, $p = 0.0645$. Specifically, when text is neutral, the difference between negative and positive prompts is significant, $p = 0.0173$. These results suggest that the two factors show an interaction pattern when text type is neutral (Figure 3b).

Different from the MLC index reported above, which does not show an interaction effect between prompt and text types, the emotion effect on the CP/C index seems to be more sensitive when the prompts are not positively affective. Overall, Chinese college EFL writers tended to use more coordinate clauses in composition when they wrote unaffectedly. The other two indices that measure the amount of coordination, namely, CP/T and T/S, do not reach a significant level in the mixed models. Similarly, none of the indices that measure subordination (C/T, CT/T, DC/C & DC/T), or degree of phrasal sophistication (CN/C, CN/T & VP/T), is significant in the analysis. The overall complexity index, C/S, is also insignificant when checked by the mixed-effects model.

Quantitative structure of syntactic complexity indices

Most of the pairwise Pearson correlations between the fourteen syntactic complexity indices are significant, except for six pairs, see Figure 4: CN/C ~ VP/T, $r = 0.01$, $p = 0.46$; CN/T ~ CP/C, $r = 0.02$, $p = 0.22$; CP/C ~ MLS, $r = 0.03$, $p = 0.18$; T/S ~ VP/T, $r = 0.03$, $p = 0.08$; T/S ~ C/T, $r = 0.02$, $p = 0.25$; T/S ~ DC/T, $r = 0.01$, $p = 0.54$; and MLC ~ MLS, $r = 0.02$, $p = 0.39$. Next, a series of reliability tests reveal that the indices that were originally designed to measure in a common linguistic dimension are not necessarily consistent in the measurements. For instance, the measures of length of production unit (MLC, MLS & MLT) are not very reliable, Cronbach's $\alpha = 0.522$, 95% confidence interval (CI) = [0.489, 0.533]. For the measures of amount of subordination (C/T, CT/T, DC/C & DC/T), Cronbach's $\alpha = 0.752$, CI = [0.736, 0.767]; For the measures of amount of coordination (CP/C, CP/T & T/S), Cronbach's $\alpha = 0.037$, CI = [-0.029, 0.099]. For the measure of degree of phrasal sophistication (CN/C, CN/T & VP/T), Cronbach's $\alpha = 0.721$, CI = [0.702, 0.739]. The measure of overall sentence complexity (C/S) was defined separately, while it is found highly correlated with MLS, $r = 0.94$, $p < 0.01$. These results suggest that the original grouping method of the fourteen indices is potentially problematic from a quantitative perspective.

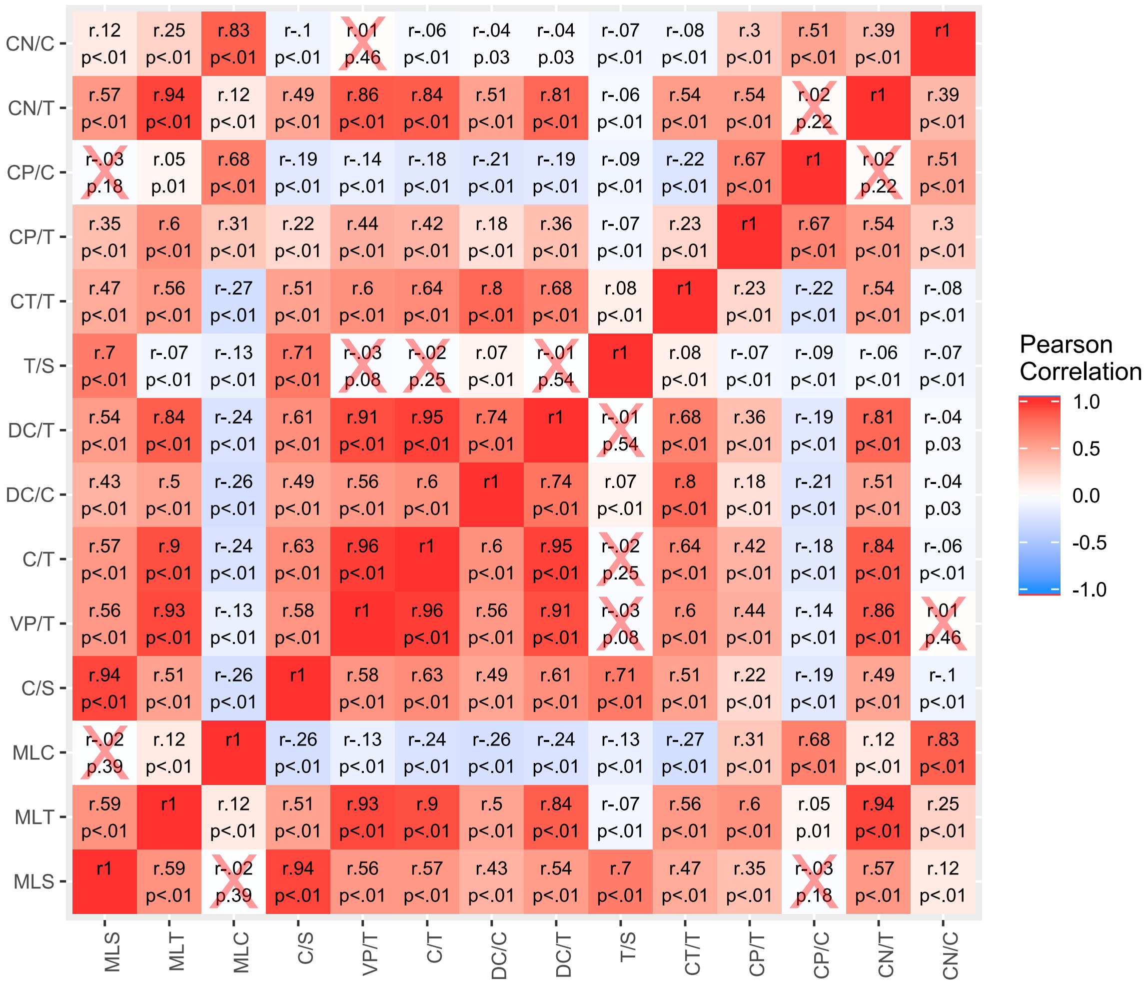

Figure 4. Correlation matrix of fourteen syntactic complexity indices based on the corpus using the Pearson method
Note: Nonsignificant coefficients are marked by a red X

To find an alternative grouping method, a principal components analysis (PCA) was conducted, and the yielded model classifies the fourteen indices into five groups (principal components, PCs) using *oblimin* rotation (Table 2). The model is significant when checked by a Chi-square test, χ^2 (41) = 14571.093, p < 0.001. Between the five components, the highest correlation is found between PC1 and PC3, r = 0.605, while other correlations are rather weak (< 0.5), suggesting that the analysis yielded a robust model. In particular, PC1 consists of five indices that measure the syntactic complexity at the T-unit level (C/T, CN/T, DC/T, MLT & VP/T). PC2 contains three indices that measure sentence-level complexity (C/S, MLS & T/S), while PC3 contains two subordination indices (CT/T & DC/C). PC4 seems to measure clause-level complexity since it contains two clause complexity indices (CN/C & MLC). Finally, PC5 contains two indices, CP/C & CP/T, both of which measure coordination complexity. Previously, two indices were shown to be sensitive to sentiment and emotion difference based on the corpus data, and they measure complexity at rather high syntactic levels, i.e., clause coordination (CP/C), and clause overall (MLC).

Table 2. Component loadings in the principal components analysis (PCA) for the fourteen syntactic complexity indices (using *oblimin* rotation*)*

Index	PC1	PC2	PC3	PC4	PC5	Uniqueness	Comment
C/T	0.917					0.023	T-unit level indices
CN/T	0.885					0.06	
DC/T	0.737					0.07	
MLT	0.931					0.027	
VP/T	0.953					0.045	
C/S		0.792				0.03	Sentence level indices
MLS		0.818				0.03	
T/S		1.01				0.036	
CT/T			0.919			0.12	Subordination indices
DC/C			0.994			0.081	
CN/C				0.995		0.046	Clause level indices*
MLC*				0.839		0.081	
CP/C*					0.874	0.053	Coordination indices*
CP/T					0.902	0.052	

* Significant indices in the sentiment analysis

General discussion

Emotion in writing prompts and written texts

The regression analysis revealed that the emotionality of writing topics is a significant predictor of the emotional score for the L2 writers' production. It is observed that the two scores are positively correlated, as students tend to write more positively when the topic itself excites more positive moods. However, the two emotional scores have different mean values, and L2 writers' production responses were close to neutral (sentiment score ≈ 0), when the prompts were the most negative ones (sentiment score ≈ -1), see Figure 1. This result suggests that among Chinese EFL writers the writing prompts and responses have different baseline sentiment values. Future research should investigate whether this pattern is specific for EFL learners of Chinese backgrounds, given the evidence from previous studies that sociocultural factors can affect L2 writers' performance (Slavkov, 2015; Storch, 2018). Alternatively, this difference may reflect the different stylistic features of writing prompts and responses. The findings of the present study are consistent with Clachar (1999), who analyzed cognitive distribution during L2 writing based on the emotionality of the prompts alone. However, the method used in the present study does not assume a deterministic link between the prompts and the writing responses in terms of emotional states. In line with previous studies on individual differences in L2 writing (Andringa, Olsthoorn, van Beuningen, Schoonen, & Hulstijn, 2012; Kormos, 2012; Llanes, Tragant, & Serrano, 2018), the present analysis observed variation in individualistic sensitivity towards the emotional information in the writing prompt, along with the general tendency, as suggested by a significant but weak positive correlation, $r = 0.351$, $p < 0.001$.

Emotionality and syntactic complexity in L2 writing

There is significant evidence in the current study that Chinese college-level EFL writers produced at different levels of syntactic complexity as a function of both the emotionality of writing prompts and the emotionality of their writing. However, not all of the fourteen L2SCA indices (Lu, 2010) are significant in the current analysis, which suggests that the indices may have different detection sensitivities for the emotional effect and require a larger sample size to obtain the statistical power. For the two significant indices (MLC and CP/C), the highest values were found when the written texts were judged as neutral, while both positive and negative writing led to a lower score (see Figure 3). This pattern seems to suggest that neutral writing might be the optimal mode to assess syntactic complexity in L2 writing, and emotional perturbations during writing may hinder L2 writers' performance. This finding is again consistent with those of Clachar (1999), who observed that L2 writers spent more time planning lexical and morphosyntactic structures when given emotional topics. The present study extends Clachar's finding that both negative and positive emotions can put extra cognitive load on L2 writers' composition. The present study is also consistent with the work of Yang et al. (2015), who found that local-level complexity features differ from global-level indices when measuring across different writing prompts. It is worth noticing that both MLC (mean length of clauses) and CP/C (coordinate phrases per clause) are indices that measure complexity at the clausal level, which is between the local (e.g., T-unit) and the global (e.g., sentence) levels. Future research may further investigate the role of emotion in syntactic processing at different linguistic levels during L2 writing.

Quantitative structure of the L2SCA indices

Quantitatively, the present study has found that the original grouping of the fourteen indices might be problematic. Evidence from the reliability tests shows that the measures do not correlate well with other indices in the same group, especially for the three indices (CP/C, CP/T & T/S) that were grouped to measure the amount of coordination, as suggested by a Cronbach's α value that is close to zero. It seems that the original grouping method is more qualitative than quantitative, but since the particular measures can be put in either the nominator (e.g., C in C/T) or the denominator (C in CP/C), it is hard to avoid arbitrary decisions in assigning a particular complexity index to any group based on purely qualitative analysis. One possible way to improve consistency is to introduce a variety of new indices (e.g., Yang et al., 2015). On the other hand, the PCA analysis in the current study divided the fourteen indices into five groups based on purely quantitative similarities. The PCA results suggest that a common denominator may group some indices, e.g., PC1 contains C/T, CN/T, DC/T and VP/T, which measure complexity at the T-unit level; while other indices might be grouped based on a shared nominator, e.g., PC5 consists of CP/C, and CP/T, both of which can measure coordination complexity in the text. The correlation matrix reveals that not all the complexity indices are positively correlated (if significant). Future research is needed to investigate the choice of complexity indices and how the selected indices can be grouped in a more defensible way.

Conclusion

The present study offers evidence that text-based methods alone are sufficient for an investigation on both the psychological status and writing performance of L2 learners. Emotion plays a vital role in language learning and acquisition, but it can also lead to potential risks in educational assessment and other testing issues. For instance, emotional writing prompts may put excessive cognitive pressure on EFL writers, which may lead them to write syntactically simpler clauses and sentences. This may cause underestimation of the EFL writers' writing proficiency. Comparability between writing scores from prompts that have different emotional polarities is at potential risks, and assessment adopting certain testing methods may need further considerations for justification. However, some limitations should be noted. First, although the selected corpus is large, essays from the TECCL corpus were written in different educational scenarios, e.g., examination writing might differ qualitatively from essays written for homework. Second, the emotional information was measured only using the text-based approach, which simplifies human emotion on a one-dimensional scale. More accurate control methods (e.g., psycholinguistic experiments) may improve the accuracy in detecting emotions during L2 writing. Finally, other potentially relevant variables such as genre, and language proficiency level were not included in the analysis. Future research is needed to investigate whether other factors can interact with the emotional variable and exert different influences on the syntactic complexity of EFL writing.

Overall, the results of the present study have revealed important and intricate relationships between emotion and writing in an EFL setting, which should be taken into consideration when teaching and testing L2 writing in different educational scenarios. As Clachar (1999) points out, practitioners should emphasize "a continuum of topics whereby aspects of linguistic and strategic knowledge as well as the various mental processes associated with writing must eventually be interconnected and manipulated by the L2 writer to function as an integral whole" (p. 57). The findings of the present study further encourage language teachers and testers to be highly aware of the role of sentimentality when conducting L2 writing tasks, and they should also be careful in choosing writing prompts for high-stake tests in order to avoid construct-irrelevant interferences.

About the author

Yizhou Wang is a PhD candidate at the School of Languages and Linguistics, the University of Melbourne, Australia. He holds a degree of Master of Applied Linguistics from the University of Melbourne, and his research interests include applied psycholinguistics, second language acquisition, and speech science. He can be contacted at: yizwang3@unimelb.edu.au

References

Andringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., & Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: An individual differences approach. *Language Learning*, *62*, 49–78. https://doi.org/10.1111/j.1467-9922.2012.00706.x

Bach, D. R., & Dayan, P. (2017). Algorithms for survival: a comparative perspective on emotions. *Nature Reviews Neuroscience*, *18*(5), 311–319.

Balahur, A., Mihalcea, R., & Montoyo, A. (2014). Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications. *Computer Speech & Language*, *28*(1), 1–6. https://doi.org/10.1016/j.csl.2013.09.003

Bradley, M. M., & Lang, P. J. (2000). Measuring emotion: Behavior, feeling, and physiology. *Cognitive Neuroscience of Emotion*, *25*, 49–59.

Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., … Rosen, B. R. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron*, *17*(5), 875–887.

Bulté, B., & Housen, A. (2012). Defining and operationalizing L2 complexity. In *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21–46). Amsterdam, the Netherlands: John Benjamins.

Clachar, A. (1999). It’s not just cognition: The effect of emotion on multiple-level discourse processing in second-language writing. *Language Sciences*, *21*(1), 31–60. https://doi.org/10.1016/S0388-0001(98)00014-X

Davis, M., & Whalen, P. J. (2001). The amygdala: Vigilance and emotion. *Molecular Psychiatry*, *6*(1), 13–34.

Dewaele, J. M., & Pavlenko, A. (2002). Emotion vocabulary in interlanguage. *Language Learning*, *52*(2), 263–322.

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science*, *298*(5596), 1191–1194.

Du, X. (2009). The affective filter in second language teaching. *Asian Social Science*, *5*(8), 162–165.

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, *26*(1), 59–84. https://doi.org/10.1017/S0272263104026130

Fang, W. (2012). The development of transnational higher education in China: A comparative study of research universities and teaching universities. *Journal of Studies in International Education*, *16*(1), 5–23.

Gabryś-Barker, D. (2009). Emotion versus cognition, or what psycho and neurolinguistics tell us about affectivity in second language acquisition. In J. Arabski & A. Wojtaszek (Eds.), *Neurolinguistic and Psycholinguistic Perspectives on SLA* (pp. 44–63). Salisbury, UK: Short Run Press.

Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment*, *26*(1), 41–54.

Kaur, J., & R. Saini, J. (2014). Emotion detection and sentiment analysis in text corpus: A differential study with informal and formal writing styles. *International Journal of Computer Applications*, *101*(9), 1–9. https://doi.org/10.5120/17712-8078

Kean, D. K., Glynn, S. M., & Britton, B. K. (1987). Writing persuasive documents: The role of students’ verbal aptitude and evaluation anxiety. *The Journal of Experimental Education*, *55*(2), 95–102.

Kissler, J., Herbert, C., Winkler, I., & Junghofer, M. (2009). Emotion and attention in visual word processing -An ERP study. *Biological Psychology*, *80*(1), 75–83. https://doi.org/10.1016/j.biopsycho.2008.03.004

Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing*, *21*(4), 390–403. https://doi.org/10.1016/j.jslw.2012.09.003

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford, UK: Oxford University Press.

Kyle, K. (2016). *Measuring Syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication*. (Unpublished doctoral dissertation). Georgia State University, Atlanta, GA.

LeDoux, J. E. (1989). Cognitive—emotional interactions in the brain. *Cognition and Emotion*, *3*(4), 267–289. https://doi.org/10.1080/02699938908412709

Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge, UK: Cambridge University Press.

Llanes, À., Tragant, E., & Serrano, R. (2018). Examining the role of learning context and individual differences in gains in L2 writing performance: the case of teenagers on an intensive study-abroad programme. *The Language Learning Journal*, *46*(2), 201–216. https://doi.org/10.1080/09571736.2015.1020332

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474–496. https://doi.org/10.1075/ijcl.15.4.02lu

Lu, X. (2011). A Corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*(1), 36–62. https://doi.org/10.5054/tq.2011.240859

Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing*, *34*(4), 493–511. https://doi.org/10.1177/0265532217710675

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. https://doi.org/10.1016/j.jslw.2015.06.003

Magezi, D. A. (2015). Linear mixed-effects models for within-participant psychology experiments: an introductory tutorial and free, graphical user interface (LMMgui). *Frontiers in Psychology*, *6*, 2. https://doi.org/10.3389/fpsyg.2015.00002

Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., & Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion*, *5*(2), 175–190.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*(4), 555–578. https://doi.org/10.1093/applin/amp044

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

Pishghadam, R. (2009). A quantitative analysis of the relationship between emotional intelligence and foreign language learning. *Electronic Journal of Foreign Language Teaching*, *6*(1), 31–41.

Provoost, S., Ruwaard, J., van Breda, W., Riper, H., & Bosse, T. (2019). Validating automated sentiment analysis of online cognitive behavioral therapy patient texts: An exploratory study. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01065

R Core Team. (2019). *R: A language and environment for statistical computing*. Retrieved from http://www.r-project.org/

Schumann, J. H. (1994). Where is cognition?: Emotion and cognition in second language acquisition. *Studies in Second Language Acquisition*, *16*(2), 231–242. https://doi.org/10.1017/S0272263100012894

Schupp, H. T., Junghöfer, M., Weike, A. I., & Hamm, A. O. (2003). Attention and emotion: an ERP analysis of facilitated emotional stimulus processing. *Neuroreport*, *14*(8), 1107–1110.

Slavkov, N. (2015). Sociocultural theory, the L2 writing process, and google drive: Strange bedfellows? *TESL Canada Journal*, *32*(2), 80–94.

Stone, P. J., Bales, R. F., Namenwirth, J. Z., & Ogilvie, D. M. (1962). The general inquirer: A computer system for content analysis and retrieval based on the sentence as a unit of information. *Behavioral Science*, *7*(4), 484–498.

Storch, N. (2018). Written corrective feedback from sociocultural theoretical perspectives: A research agenda. *Language Teaching*, *51*(2), 262–277. https://doi.org/10.1017/S0261444818000034

Swain, M. (2013). The inseparability of cognition and emotion in second language learning. *Language Teaching*, *46*(2), 195–207. https://doi.org/10.1017/S0261444811000486

Taguchi, N., Crawford, W., & Wetzel, D. Z. (2013). What linguistic features are indicative of writing quality? A case of argumentative essays in a college composition program. *TESOL Quarterly*, *47*(2), 420–430.

Vaughn, C., & Leff, J. (1976). The measurement of expressed emotion in the families of psychiatric patients. *British Journal of Social and Clinical Psychology*, *15*(2), 157–165.

Vyatkina, N. (2013). Specific syntactic complexity: Developmental profiling of individuals based on an annotated learner corpus. *The Modern Language Journal*, *97*, 11–30. https://doi.org/10.1111/j.1540-4781.2012.01421.x

Way, D. P., Joiner, E. G., & Seaman, M. A. (2000). Writing in the secondary foreign language classroom: The effects of prompts and tasks on novice learners of French. *The Modern Language Journal*, *84*(2), 171–184. https://doi.org/10.1111/0026-7902.00060

Weinberg, A., & Klonsky, E. D. (2009). Measurement of emotion dysregulation in adolescents. *Psychological Assessment*, *21*(4), 616–621.

Wu, J. (2015). Distributions of scientific funding across universities and research disciplines. *Journal of Informatrics*, *9*(1), 183–196.

Xue, X. (2015). *Ten-thousand English compositions of Chinese learners (the TECCL corpus) (version 1.1)*. The National Research Centre for Foreign Language Education, Beijing Foreign Studies University: http://corpus.bfsu.edu.cn/content/teccl-corpus

Yang, W., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*, *28*, 53–67. https://doi.org/10.1016/j.jslw.2015.02.002

Appendix. Type III ANOVA results for linear mixed models of each syntactic complexity index

Index	Factor	Sum Sq.	Mean Sq.	NumDF	DenDF	F-value	P-value	Sig.
MLC	Prompt	58.132	29.066	2	2611	3.5008	0.0302	*
	Text	85.755	42.877	2	2611	5.1642	0.0057	**
	Interaction	33.635	8.409	4	2611	1.0128	0.3994	n.s.
MLS	Prompt	341.46	170.73	2	2610.9	0.4815	0.6179	n.s.
	Text	1118.8	559.42	3	2605.7	1.5778	0.2066	n.s.
	Interaction	55.33	13.83	4	2610.3	0.0390	0.9971	n.s.
MLT	Prompt	324.64	162.32	2	2611	1.1717	0.3100	n.s.
	Text	252.25	126.12	2	2611	0.9104	0.4025	n.s.
	Interaction	417.46	104.36	4	2611	0.7534	0.5557	n.s.
C/T	Prompt	2.4670	1.2335	2	2611	0.6171	0.5396	n.s.
	Text	2.7945	1.3973	2	2611	0.6991	0.4971	n.s.
	Interaction	4.7971	1.1993	4	2611	0.6000	0.6627	n.s.
CT/T	Prompt	0.0073	0.0037	2	2586.8	0.0636	0.9384	n.s.
	Text	0.1259	0.0629	2	2609.2	1.0941	0.3350	n.s.
	Interaction	0.13152	0.0329	4	2604.4	0.5717	0.6832	n.s.
DC/C	Prompt	0.0231	0.0116	2	2610.9	0.5222	0.5933	n.s.
	Text	0.0653	0.0326	2	2605.1	1.4751	0.2290	n.s.
	Interaction	0.0395	0.0099	4	2610.1	0.4459	0.7754	n.s.
DC/T	Prompt	1.4292	0.7146	2	2611	0.7089	0.4923	n.s.
	Text	2.6671	1.3335	2	2611	1.3229	0.2666	n.s.
	Interaction	2.7595	0.6899	4	2611	0.6843	0.6028	n.s.

CP/C	Prompt	0.2929	0.1465	2	2601.9	4.7739	0.0085	**
	Text	0.2998	0.1499	2	2580.0	4.8865	0.0076	**
	Interaction	0.3622	0.0905	4	2607.2	2.9513	0.0190	*
CP/T	Prompt	0.5833	0.2916	2	2604.7	2.0164	0.1333	n.s.
	Text	0.7354	0.3677	2	2610.4	2.5423	0.0789	n.s.
	Interaction	0.6937	0.1734	4	2608.8	1.1990	0.3091	n.s.
T/S	Prompt	0.1278	0.0639	2	2610.9	0.0956	0.9088	n.s.
	Text	0.7954	0.3977	2	2609.3	0.5954	0.5514	n.s.
	Interaction	0.9225	0.2306	4	2609.9	0.3453	0.8475	n.s.
CN/C	Prompt	1.2256	0.6128	2	2589.4	2.8040	0.0608	n.s.
	Text	0.5088	0.2544	2	2609.3	1.1640	0.3124	n.s.
	Interaction	0.8011	0.2003	4	2605.3	0.9164	0.4533	n.s.
CN/T	Prompt	6.5149	3.2575	2	2611	0.9566	0.3843	n.s.
	Text	11.9276	5.9638	2	2611	1.7514	0.1737	n.s.
	Interaction	8.5064	2.1266	4	2611	0.6245	0.6450	n.s.
VP/T	Prompt	5.0898	2.5449	2	2611	0.6643	0.5147	n.s.
	Text	4.6253	2.3127	2	2611	0.6037	0.5469	n.s.
	Interaction	5.3490	1.3373	4	2611	0.3491	0.8448	n.s.
C/S	Prompt	2.9967	1.4984	2	2611	0.2785	0.7569	n.s.
	Text	8.2061	4.1030	2	2611	0.7627	0.4665	n.s.
	Interaction	0.3803	0.0951	4	2611	0.0177	0.9994	n.s.

* Significance level: $p < 0.5$
** Significance level: $p < 0.01$