ATLANTIS
PRESS

A Textual Analysis of Post-reading Writings in Light of the Alignment Effect and Its Implications on Teaching Writing to EFL Learners

Suyun Xu

Xi'an Jiaotong University City College, No. 8715 Shangji Rd., Weiyang, Xi'an, China

Abstract. As is suggested by the Alignment Effect, EFL writings are featured by remarkable textual qualities when they are finished after close-reading of model texts. With textual analyzers, this research collects data of lexical complexity, syntactic complexity and textual coherence of 126 pieces of post-reading writings by EFL. After the analysis, the research finds simple words dominate while complex content words rarely occur. Content verbs, adjectives, and adverbials show much less diversity than nouns. The EFL writers produce full and long sentences which are mainly compound and simple, but need improvements in composing imbedded dependent clauses and complex noun phrases. Due to connectives abuse and loose semantics, sentences within paragraphs are poorly connected, while paragraphs run smooth from one to another. The textual features of post-reading writings yield some pedagogical references to the teaching of writing to EFL learners.

Introduction

English writing is a comprehensive output activity in which learners express their thoughts and ideas with language structures and norms acquired by intensive reading of words, sentences and discourses (widely known as language inputs). Just as effective language outputs are based on abundant and intensified inputs, smooth writings should be preceded by close in-depth reading. But reading does not necessarily guarantee qualified writings. Then how does reading produce an impact on writing? This study attempts to unveil the textual features of writings which are completed after close reading of model texts, and discusses what implications those features will yield on teaching writing to EFL learners.

Theories

Understanding the Alignment Effect. In recent years, the Alignment Effect in language learning has attracted more and more researchers of second language acquisition (SLA). In the Alignment Model of Pickering & Garrod (2004), speakers tend to adjust and adapt the conversation to each other until they reach some agreements in language patterns and meanings. Those agreements are generally witnessed as repeating and imitating the other's pronunciation, words, and sentences. In reaching the agreements, language outputs are closely tied to inputs, which is known as "structural coordination". The structural coordination enables language learners to acquire syntactic patterns (Wang, 2012) and apply them in speaking and writing. Based on the Model, he proposes a road path to successful language learning: interaction → comprehension → coordination → outputting → acquisition. He believes comprehension is a powerful potent engine to make for one's language achievement as comprehension always goes before the output. In writing, learners apply the language patterns acquired in reading, and try to make their written texts to be identical to the model ones. Then the distance between inputs and outputs is narrowed, and the weaker writing is gradually lifted up to the stronger reading. Therefore, the key step toward effective language learning is to combine writing with reading and generate the Alignment Effect in their coordination.

Many researches (Wang, 2015; Yang, 2015; Jiang & Tu, 2016) affirm the Effect when learners try to imitate model language patterns of the input in their communicative output. Meanwhile, the Effect can reinforce and strengthen the model language, and encourage learners to timely use the model patterns and expressions acquired from the reading task. Thus the learners get improvements

in language learning. The Alignment Effect is also effective in developing speaking and translation (Xu, 2016).

Measuring English Writing. The afore-mentioned researches have affirmed the Alignment Effect positively, but cannot exactly measure the quality of language outputs. They just score them globally, without fully specifying the textual features of the outputs. Even thought they do examine language details, they resort to exemplification instead of quantitative methods. Since there isn't any smart instrument to extract both surface and deep features of written texts, those researches engaged in EFL writing fail to deal with scaled data when working on the potent items which takes time and arduous brain labors (Lu & Xu, 2016).

EFL writings are scored mainly with their words, sentences and discourse coherence. The instructor of EFL writings should clearly know to what extend learners have achieved lexically, syntactically, and textually, and then take measures accordingly. Researchers have developed some measurements of lexical, syntactical, and textual features. At the lexical level, word variety, density and complexity are commonly examined. A syntactical study usually examines the subordinate or coordinate structure, sentence complexity or variety, and phrasal expansion (Norris & Ortega, 2009). At the textual level, coherence is primarily valued, but it is far from enough to be assessed only by cohesive devices. Indeed, a text should be assessed by such potent factors as coherence across sentences and paragraphs, and the potent semantic relations.

The Textual Measurement System. EFL writings vary with different learners, tasks, and teaching methods, and are featured by the textual features of words, sentences and coherence. The intermediate and advance learners are instructed to use proper collocations, adapt to the context, and choose appropriate words. As for the syntactic, they are taught strict grammatical rules and various sentence patterns before they can adeptly use nonfinite verbs, gerunds, and prepositions. At the discourse level, they should pay special attention to the controlling idea, logics, transition, and even rhetoric and style. A piece of writing should be evaluated not only on its holistic layout, but also on its detailed expressions and idea development.

Lexical Complexity and Its Measuring Items. Words are indispensable for language learning, and also the building blocks of EFL writing. Malvern et al. (2004) defines lexical complexity into three dimensions: density, difficulty, and variety. Lexical density is referred as the ratio of concrete words which include nouns, verbs, adjectives and adverbials. Lexical density can be gained by counting average length of a sentence, and the ratio of concrete words to the total words. Lexical difficulty can be gained by counting the type and token of advanced concrete words in a text. Lexical diversity means that synonyms, near-synonyms, and even antonyms are employed to avoid sheer repetition. Sometimes changing the formation can also add to the diversity.

Syntactic Complexity and Its Measuring Items. Sentences are the core of writing, connecting individual words and developing into a full text. Syntactic complexity means that sentences are compound, complex and various enough. It is usually regarded as an indicator of one's writing proficiency, bearing a close relation with writing quality. It is generally taken as a major element in assessing language proficiency. According to Norris & Ortega (2009), syntactic complexity should be measured by five dimensions: dependent clauses, overall complexity, phrasal expansion, compound structure, and sentence pattern variety. Sentence, T-unit, and clause are the major component to measure a learner's syntactic development. Those indexes of syntactic complexity measure the length of a sentence, length of a T-unit, number of clauses, the ratio of complex T-units, number of coordinate phrases, etc.

Textual Coherence and Its Measuring Items. Textual coherence involves far more than cohesive devices, as a coherent text does not necessarily employ many cohesive devices and more devices do not necessarily make for a coherent text. (McNamera et al., 2010). The study of textual coherence occurs at a local and the overall part. Local coherence is realized by such cohesive devices as coordination, substitution, and connection which can be measured by the indexes of argument overlap and adjacent semantic. Overall coherence is achieved when all sentences of a paragraph are centered on one controlling idea, and all paragraphs are focused on one topic. Since it is potent, it is

hard to be observed. But such indexes as stem overlap and latent semantic analysis (LSA) can reveal it.

Research Methods

Influenced by the Alignment Effect, language learners consciously or unconsciously use the words, expressions, and sentence patterns acquired through intensive reading when they write upon a similar topic. This study examines some post-reading writings, and tries to discover their textual features which are expected to yield quantitative supports for EFL writing.

By quantitatively exploring the textual features of post-reading writings, this study aims to respond to the question: What are the lexical complexity, syntactic complexity, and textual coherence of post-reading writings respectively?

Subjects. The researcher followed the course of "Intermediate Intensive Reading" to sophomores of English major at a college that is averagely listed in China. In the one-semester course, the students carefully studied six units each of which contained a reading text by English native professional writers and some supplementary exercises. They carefully study the usage of new and key words, expressions, patterns, and worked out the text structure. After immerging into the language knowledge, they were given the post-reading writing task which was closely related with the text topic but adjusted to the young students or the social currency. Then the researcher input the 126 pieces of writings (6 each by 21 students) into a mini corpus in the computer.

Instruments. This study employed Lexical Complexity Analyzer (LCA) by Lu Xiaofei and Ai Haiyang of Pennsylvania University to get data of lexical complexity, and their Second Language Syntactic Complexity Analyzer (L2SCA) to for syntactic complexity. Both LCA and L2SCA are smart analyzers based on computational linguistics and Stanford parse. They are specially adapted to ESL/EFL learners. Simple but practical, they can deal with single and batch texts. L2SCA's accuracy in identifying sentence structure is 0.83-1.0 (Lu, 2010), and both are significantly related to manual analysis by 0.834-1.0 (Yoon & Polio, 2016). LCA outputs both complexity and diversity data. The basic count concerns token, type, and TTR. L2SCA, focused on sentence, T-unit, and clause, can examine sentence complexity, T-unit complexity, coordinate and dependent clauses, and phrases and non-finites as well (Yang, Lu, & Weigle, 2015).

For textual coherence, this study adopted Coh-Metrix 3.0 by McNamera of Memphis University. It can reveal both surface and latent features of a text with 11 types of indexes. This study just chose four out of them: Text Easability, Connectives, Referential Cohesion, and Latent Semantic Analysis.

Data Collection and Analysis. Two college English teachers were invited to act as the scorers to grade the 126 pieces of writings. Both of them have rich experience for more than eight years of grading writings for College English Test (Band 4), a nationally recognized English proficiency test in China which is universally taken by Chinese college students. They gave points (15 fully) to each piece independently, and the average point of each writing was taken as its score.

LCA, L2SCA, and Coh-Metrix were employed to collect quantitative data of lexical, syntactical and textual features of all the writings. Those basic data were then analyzed by SPSS for description. With scores as the dependent variable, and the basic data as the independent variables, SPSS made correlation and regression analysis in order to reveal the contribution of words, sentences and coherence to the quality of those writings.

Results and Discussion

Lexical Complexity Analysis. The lexical data are analyzed by SPSS for descriptive statistics and their correlation with scores, and the results are shown in Table 1. As Test for English Majors states the writing should contain about 200 words, the post-reading writings of this study averagely contain 216.5238 words. Therefore, those writings meet up the length requirement for English majors who generally take the Test in the second year of college study. The scores range from 6 to 14 points, with a mean of 10.88889, and a scoring average of 72.59%.

Table 1 Lexical Results

Type	Lex. Index	Mean	Correlation with Scores	
			Pearson	Sig. (2-tailed)
Length	W	**216.5238**	**.497****	**.000**
	LD	**.5154**	.112	.213
Difficulty	LS1	**.2454**	.207*	.020
	LS2	**.1930**	**.239****	**.007**
	VS1	**.1497**	.190*	.033
Diversity	NDW	114.6190	**.542****	**.000**
	NDWERZ	38.9040	.200*	.025
	TTR	.5382	-.163	.068
	MSTTR	.7689	.219*	.014
Variety	VV	.2027	-.085	.342
	NV	.6619	-.149	.096
	ADJV	.1314	.010	.909
	ADVV	.1003	-.016	.855

Concrete words describe a subject and state one's viewpoints more accurately than function and form words. In the post-reading writings, concrete words just account for a half of the full length (LD=.5154), lower than .5912 of their peers of key universities (Zhu & Wang, 2013). Sophisticate concrete tokens' ratio to all concrete words is .2454, and sophisticate concrete types' ratio to word types is .1930. By diversity, the mean of non-repetitive tokens (including concrete, function and form words) is 114.6190, slightly over half of total tokens. The diversity mainly comes from noun words (NV=.6619), but verbs, adjectives, and adverbials are less various.

Among the lexical indexes, W (.497**), LS2 (.239**), and NDW (.542**) are significantly related to scores. The linear regression analysis with W, LS2, and NDW as the independent variables and scores as the dependent yields a writing quality predicting model from a lexical perspective. The model (F=17.045，Sig.=.000a) can predict 29.5% of the score (R2=.295). Specifically, only NDW can efficiently predict the score (Std. Co=.602, t=2.725, Sig.=.007).

Judged by the lexical results, the subjects can use concrete words to express their viewpoints in the post-reading writing task, but the concrete words are mainly common words, with less advanced ones. The concrete adjectives and adverbials are especially rare occurrence. It should be noted that words make some contribution to the quality of writing. Therefore, in writing, learners should try to adopt more words. More sophisticate and more advanced the words are, the higher quality the writing would has. Additionally, they should avoid repetitive words, and gain diversity with pronouns, synonyms, super-ordinates, hyponyms and even antonyms.

Syntactic Complexity Analysis. The descriptive and correlation results of syntactic complexity are shown in Table 2. The mean length of a sentence is of 14.977 words or of 1.102 T-units. Each T-unit is of 13.672 words or of 1.631 clauses. And each clause contains 8.448 words. MLT, an index of syntactic maturity, indicate that the subjects have obvious syntactic awareness and can produce complete T-units. C/S, the prominent index of syntactic complexity, is 1.795, indicating that the sentences are a bit sophisticated with more than one clauses in each sentence. However, most of the clauses are coordinate instead of subordinate, as the index of dependent clauses is much small (DC/C=.360). Moreover, the coordinate clauses are short and simple, just with "one subject + one predicate + one object", rather than with multi subjects or predicate or object (CP/C=.229). The dominance of simple coordinate clauses greatly erodes the syntactic complexity. The simplicity is further affirmed by CN/C (.903), which shows complex nouns in one clause is less than one while nouns function as subjects, objects, or attributives and therefore should be much more than one or even two. The low CN/C also implies the syntactic grammar is simple and lacks modifiers.

Table 2 Syntactic Results

Type	Syn. Index	Mean	Correlation with Scores	
			Pearson	Sig. (2-tailed)
Length	MLS	**14.977**	**.289****	.001
	MLT	**13.672**	**.363****	.000
	MLC	**8.448**	**.279****	.002
Complexity	C/S	1.795	.062	.490
Dependent	C/T	1.631	.163	.068
clause	CT/T	.444	.117	.193
	DC/C	.360	.154	.085
	DC/T	.611	.174	.052
Coordination	T/S	**1.102**	**-.098**	.273
	CP/T	**.364**	**.356****	.000
	CP/C	**.229**	**.289****	.001
Phrases	CN/T	**1.456**	**.349****	.000
	CN/C	**.903**	**.281****	.001
	VP/T	2.206	.171	.055

Those which are significantly related to the scores are MLS (.289**), MLT (.363**), MLC (.279**), CP/T (.356**), CP/C (.289**), CN/T (.349**), CN/C (.281**). Taking them as the independent variables, and the score as the dependent variable, the research develops a predicting model of writing quality from a syntactic perspective. The model is of a little significance (F=3.687，Sig.=.001a), only predicting 17.9% of the score ($R2$=.179). And within the model, all the seven variables do not bear a predicative co-efficiency, T value, and Sig. The results indicate a weak relation between syntactic complexity and writing quality. This might be explained by the students' general writing proficiency. The student under this study don't achieve remarkably in English. Most of their sentences are with short coordinate clauses or simple dependent clauses, unlike that of advanced learners good at complex and compound clauses and sophisticate coordinate phrases. Therefore, teachers should direct students how to seek logic relations between simple sentences and transform them into embedded attributive clauses or adverbial clauses. For advancement students can turn clauses into sophisticate noun phrases or preposition phrases.

When compared with their American peers, the students under this study obviously lag behind the former who, in Lu & Ai's (2015) research of 200 pieces of writings by American native English college students, write sentences and T-units and clauses of more length (MLS=19.602, MLT=17.308, MLC=10.692). The biggest distance between them is sophisticated noun phrases with higher CP/C (1.222) and CN/T (2.087) by American students. So EFL learners should also be instructed to produce more complex sentences by expanding the use of various sophisticate phrases.

Textual Coherence Analysis. Textual coherence derives from both surface cohesive devices and latent semantic connections between sentences and paragraphs. The coherence data are analysis by SPSS and yield the result in Table 3. The local text (7.69%) is much less coherent than the deep semantic cohesion (75.58%) of the whole text. And the text is at low level with high readability (68.07%).

Table 3 Textual Coherence Results

Type	Coh. Index	Mean	Correlation with Scores	
			Pearson	Sig. (2-tailed)
Easability	PCCONN	7.69	**-.255****	.004
	PCREF	42.27	-.097	.282
	PCDC	75.58	.008	.926
	RDFRE	68.07	**-.328****	.000
Connectives	CNCCaus	33.78	-.097	.280
	CNCLogic	45.38	-.159	.076
	CNCAdd	49.90	.170	.058
	CNCTemp	19.04	.048	.592
Reference	CRFAO1	.5282	-.024	.791
	CRFAOa	.4383	-.026	.772
	CRFCWO1	.1219	-.159	.075
	CRFCWOa	.0982	-.126	.159
Latent Semantic	LSASS1	.1671	**.205***	.021
	LSASSp	.1520	.121	.175
	LSAPP1	.2910	**.257****	.004

The local text's low coherence can be attributed to the improper use of connectives. The frequency of connectives is not low, with mean of cause, logic and addition being over 30. But the time connectives are only 19.04, mainly because the writings are all arguments which don't involve many timers. The high frequency of connectives does not add to textual coherence due to the fact that the students misuse some connectives, and sometimes abuse them. Liang (2006) shares the same findings. So teachers are expected to explain rules of connectives, and assign special exercises of clause combination, separation, and error-correction, so that students have strong awareness of connectives. Another factor related to local coherence is reference. No matter adjacent or overall sentences, they overlap moderately. It means the students are able to use pronouns, super-synonyms, and key words to join those sentences. But references are not correlated with scores. It might be because scorers' ignorance of referential cohesion.

As for latent semantic analysis, the LSA of adjacent sentences (.1671) and all sentences in a paragraph (.1520) are a bit low, indicating the students fail to observe the paragraph principles. In the post-reading writing, the learners haven't a strong awareness that a paragraph is controlled by a topic sentence. Some sentences are loose or far from the controlling idea of a paragraph and therefore can be excluded. However, the LSA of all paragraphs are high (.2910). It shows the students strictly keep the topic.

Those indexes significantly correlated with scores are PCCONN (-.255**), RDFRE (-.328**), LSASS1 (.205*), LSAPP1 (.257**). With them, a predicting model from a cohesive perspective is developed. The model is with slight significance (F=7.250，Sig.=.000a), and can predict 19.3% of scores (R^2=.193). But the four indexes are predictable independently.

Conclusion

Enlightened by the Alignment Effect, this research follows the post-reading writing task of English sophomores. It extracts quantitative features of the words, sentences, and coherence of the writings, and then makes descriptive, correlation, and regression analysis. The research gains some interesting findings. Lexically, the students employ required words, half of which are concrete ones but few are sophisticated; they use various nouns, but less changeable verbs, adjectives, and adverbials. In syntactic sense, their sentences are complete and long with coordinate clauses or phrases, but less complex as embedded subordinate structures and complex noun phrases do not commonly occur. As for textual coherence, they can keep paragraphs smooth and connected and

join clauses with referential words, but sentences within a paragraph are not focused as desired. Meanwhile, the connectives are misused and abused. Exploring textual features of EFL writings are of great pedagogical significance for improving writing and reading courses to Chinese college English writers. Some flaws with the research methods are also detected. If the post-reading writing were compared with other types of writing, the research might find how the Alignment Effect works on the progress of writing quality.

Reference

[1] Lu Xiaofei, Ai, Haiyang. 2015. Syntactic complexity in college-level English writing: differences among writers with diverse L1 background [J]. Journal of Second Language Writing, 29: 16-27.

[2] Lu, Xiaofei. 2010. Automatic analysis of syntactic complexity in second language writing [J]. International Journal of Corpus Linguistic, 15(4): 474-496.

[3] Malvern, D., B. Richards, N. Chipere , & P. Duran. 2004. Lexical Diversity and Language Development: Quantification and Assessment. Houndmills, England: Palgrave MacMillan.

[4] McNamara, D. S., S. A. Crossley & P. M. McCarthy. 2010. Linguistic features of writing quality [J]. Written Communication, 27: 57-86.

[5] Norris, J. M., & L., Ortega, 2009. Towards an organic approach to investigating CAF in instructed SLA: The case of complexity [J]. Applied Linguistic, 30: 555-578.

[6] Pickering, M. J. & S. Garrod. 2004. Toward a mechanistic psychology of dialogue [J]. Behavioral and Brain Sciences 27(2): 169-226.

[7] Yang, W., Lu X., Weigle, S.C. 2015. Different topics, different discourse: relationship among writing topic, measures of syntactic complexity, and judgments of writing quality. Journal of Second Language Writing, 28: 53-67.

[8] Yoon, H.J., Polio, C. 2016. The linguistic development of students of English as a second language in two written genres. TESOL Quarterly.

[9] Jiang Lin, Tu Mengwei. 2016. Effects of the continuation task on L2 vocabulary learning [J]. Modern Foreign Languages, 6: 819-829.

[10]Lu Xiaofei, Xu Qi. 2016. L2 syntactic complexity analyzer and its applications in L2 writing research [J]. Foreign Language Teaching and Research, 3: 409-420.