Complexity Factors in the L2 Paraphrasing Process

Steven Acton

Northern Arizona University

Abstract

Paraphrasing and rewriting texts is a valuable skill in today's academic writing. Second language (L2) writers face an uphill battle dealing with not only cultural and institutional differences, but the complexity of the language can cause multiple problems. In this study 28 students were given a rewriting task that was prepared with various levels of lexical and grammatical complexities. The copied strings were totaled, and each text rated for content accuracy. There was a wide range of amount of copied strings and the content accuracy found in paraphrases between writers and even from text to text. However, overall there was more copying with higher complexity and lower content accuracy. The findings also suggest that lexical complexity encouraged more copied strings and grammatical complexity caused lower content accuracy.

Complexity Factors in the L2 Paraphrasing Process

Background

Second language (L2) writers are expected to use external sources in their academic writing just as native writers. As they navigate through the various requirements brought on by teachers, educators, or employers, the L2 writer must also deal with an array of conflicting information that complicate the task of rewriting a text. Much discussion has been made about the confusion that may come from culture, academic setting, different expectations among disciplines, and varying definitions of paraphrasing (Keck, 2006; Pennycook, 1996; Pecorari & Shaw, 2012; Polio & Shi, 2012; Shi, 2012). While all of these complications bring to light some confusion that may plague L2 writers, the task of rewriting a given text brings its own difficulties and issues that are inherent in second language writing.

The lexio-grammatical complexity of a text may be the most significant burden placed on an L2 writer when completing a paraphrasing task. Rewriting a text places a writer in the position of trying to understand the written thoughts of another person. The words and grammar are chosen by the original writer for a specific context and background that may not be known by the individual attempting the paraphrase. The words, collocations, and noun-noun phrases, and sentence structures may have never been seen or used by the paraphrasing writer. Possibly worse may be if they have experience with one meaning of a word, but the text uses it in a different meaning. Past researchers have claimed that rewriting texts demands a higher level of cognitive thinking that requires a variety of skills, such as utilizing synonyms, reformulating sentence structure, and switching between passive and active voice (Barks & Watts, 2001; Pennycook, 1996). The L2 writer carries the weight of a limited lexicon, grammar repository, and an

insufficient reading comprehension. These deficiencies may be the instigators behind L2 writers' failures in rewriting texts.

In order to better identify the relationship between lexico-grammatical complexity and L2 paraphrasing, the current study looked at how writers would attempt to handle a paraphrasing task with a variety of texts. Previous research does not consider just how L2 writers approach lexical and grammatical complexity. Little investigation has been done into whether writers prefer to copy or attempt paraphrasing when faced with various complexities. To understand more about the process that L2 writers go through, the following research questions were posed:

Research Questions

1) Do elevated levels in lexical complexity result in low levels of paraphrasing success?

2) Does high grammatical complexity result in low levels of paraphrasing success?

Methods

Participants

All participants came from a University PIE in the Southwestern part of the US and were students in a level 5 class. A demographic questionnaire was prepared to collect relevant information. The average age was 20 years. There were 12 females, 14 males, and two that didn't choose. Out of the 28 participants, 27 were Chinese and one was Japanese. Most participants had been in the US for only a few months, but had studied English in a classroom for an average of 10 years. Out of the 28 participants, 22 marked that they had prior experience with paraphrasing.

Instruments

Texts were selected from a Gale, Cengage Learning book from the Introducing Issues with Opposing Viewpoints series titled, *Television*. The series was listed as appropriate for 7-10+ grade level on the Cengage website (http://solutions.cengage.com/Greenhaven/). Eight smaller

texts were selected with each containing at least one specific grammatical feature, which included Verb Compliment Clause, Subordinating Clause, Long Noun Phrase, Noun-Noun Phrase, Extraposed, Coordinating Conjunction, Relative Clause, and Modals. One longer text was chosen that included a Verb Complement Clause, a Coordinating Conjunction, a Subordinating clause, and two modals (can, must). The word count for the texts listed from text 1 to 9 were 49, 51, 28, 45, 33, 53, 35, 33, and 89.

The instrument was pilot tested with four friends of the researcher who were L2 speakers studying at the university. When reviewing the pilot results, it was determined that one of the texts contained a vocabulary word that was abnormally infrequent (quixotic) and another that was idiomatic (come to grips with), so participants were unable to gain an understanding of the meaning through dictionary use. As a result, *quixotic* was replaced with *idealistic* and *come to grips with* was replaced by *accept*. No other vocabulary was an obvious hindrance to completing the paraphrasing task.

Two versions of the instrument were created to counterbalance any effect that might result from the ordering of the texts or participant mental exhaustion. A consent form and demographic survey were printed together on one paper, and the task was printed double-sided on two other papers (see Appendix A).

Procedures

The task was administered during a weekly writing class. Participants were told that the researcher wanted to analyze how well they could rewrite the texts. How to complete the task and the directions were explained and then about one hour remained of class time. The task instructions asked participants to rewrite the texts in their own words and sentence structure.

Dictionary use was also allowed. As participants finished the task, they submitted their work and left the classroom.

Data Analysis

The lexical complexity was checked using AntWordProfiler (Anthony, 2013). General statistics were calculated for each level of the word files. The vocabulary frequency files based on Paul Nations research that include the first 1,000 for level 1; 2,000 for level 2; and 570 for level 3. Level 4 included any words not within the first three levels. Token percent for level four was evaluated for degree of complexity. The texts fell into three groups: 6-7% = L, 11-12% = M, and 20-29% = H (see Table 1).

Table 1

Lexical Text Complexity Ranges

	H	M	L
Lexical Comp	20-29%	11-12%	6-7%
Texts	3, 4	5, 6, 7, 8, 9	1, 2

The grammatical complexity was calculated by counting the dependent/independent clause ratio with the lowest ration being 0 (texts three and four) and the highest being 4/1 (text seven) and dividing up the ratio scores into High (4-2) and Low (1.9-0) ranges (see Table 2).

Table 2

Grammatical Complexity Ranges

	H	L
Grammatical Complexity	4.0 -2.0	1.5-0
Texts	1, 5, 7, 8, 9	2, 3, 4, 6

Word strings that matched the original text with at least two content words (nouns, verbs, adjectives, adverbs) were marked as copied words following the example of previous

paraphrasing research (Shi, 2004; McDonough et al, 2014). Proper nouns and numbers were ignored when searching for copied strings; however, once strings were identified, any copied proper nouns or numbers in a string were included in the word count in order to accurately compare the copied strings. Furthermore, when participants copied a string but omitted a word from the original text that was necessary for grammatical correctness without replacing the word, the word was still counted as part of the copied string. A copied number in a string was counted as one word, and numbers and pronouns were not considered content words. Copied strings were counted regardless of punctuation around or in the strings. Strings of two, three to four, five to six, and seven or more were grouped together for comparison. Finally, for the analysis all strings of five or more words were considered to allow for the necessary use of some words and short phrases.

When analyzing the data, it was discovered that participant three underperformed on all paraphrasing texts with minimal effort and generally writing only one sentence for each text. Also, participants 10, 24, 25, 26, and 28 did not complete all nine paraphrasing tasks. These participants were not included in the copied strings comparison, but were part of the content accuracy comparison when scores were available. This allowed the content accuracy numbers to be based on more data observations, which would provide more confidence in the overall ratings.

In order to score the rewritten texts according to content accuracy, each text was divided into three sections of meaning. Each text was given a score from 0-3 based on how many of the sections were accurately conveyed. If the paraphrased text had section one and three sufficiently rewritten but not section two, then an accuracy score of two was given. In this way, a more consistent scoring was achieved for content accuracy. For example, text 1 was divided into the following three sections:

1) The responses she gets generally fall into one of two camps. 2) One group says personal responsibility died a slow but certain death in this country a long time ago 3) and that she is just another principled but idealistic dreamer who has yet to accept the inevitability of government censorship (*Television*).

Results

The number of copied strings for each text, including the total of all strings of five words or more, and the content accuracy descriptive statistics are shown below (see Table 3). The average percent copied for the nine texts ranged from 18% to 47%, while the content accuracy was between 1.56 to 2.59.

Table 3

Complexity Ratings and Scores for Comparison

	Text 1	Text 2	Text 3	Text 4	Text 5	Text 6	Text 7	Text 8	Text 9
Total Words	49	51	28	45	33	53	35	33	89
Lexical Comp.	L	L	H	H	M	M	M	M	M
Gram. Comp.	H	L	L	L	H	L	H	H	H
Percent Copied	0.30	0.18	0.37	0.45	0.29	0.32	0.47	0.31	0.31
Total of all 5+ n=22	32	17	22	38	15	31	32	16	62
Very High 7+	17	10	10	16	9	15	20	10	27
High 5 or 6	15	7	12	22	6	16	12	6	35
Med 3-4	19	16	20	36	18	23	9	16	28
Low 2	18	5	5	2	23	5	2	14	13
	n=26	n=27	n=27	n=27	n=28	n=28	n=28	n=25	n=25
Content Acc.	1.56	2.08	2.00	2.19	2.52	2.59	2.07	1.79	1.83
St D	1.08	0.84	0.8	0.8	0.64	0.69	0.73	0.88	0.96
Min	0	0	1	1	1	0	1	0	0
Max	3	3	3	3	3	3	3	3	3

If the texts are ranked by number of copied strings, then text 9, which has two or three times the length as the other texts, comes out with the most (see Table 4). Text 5 (15 strings) and

text 8 (16 strings) are at the bottom; however, text 2 is a close third with 17, and text 2 has a considerably higher total word count of 18 more words. Considering text 2 is rated low in both lexical and grammatical complexity, that is not a surprise. Text 5 (Lex = M, Gram = H), on the other hand, is a bit of a rogue. We would expect a higher volume of copied strings and lower content accuracy, but found the opposite. The previous conversation about the types and location of the grammatical complexity within text 5 may be the difference. Also, texts 5, 8, and 3 have the lowest total words out of all of the texts with 33, 33, and 28 respectively.

Table 4

Ranking by Copied Strings

	Lexical	Grammatical	Strings
Text 9	M	H	62
Text 4	H	L	38
Text 7	M	H	32
Text 1	L	H	32
Text 6	M	L	31
Text 3	H	L	22
Text 2	L	L	17
Text 8	M	H	16
Text 5	M	H	15

By ranking texts according to content accuracy, Text 5, and text 6 to a lesser degree, fail to follow the pattern of more complexity being more difficult to paraphrase (see Table 5). However, neither text has many noun chunks, and the grammatical complexity of text 5 is isolated enough not to greatly affect the accuracy of the paraphrases. Text 4 also shows up high on the list at three, but a closer look at the copied strings reveals that much of the high accuracy rating is due to rampant copying.

Table 5

Ranking by Content Accuracy

	Lexical	Grammatical	Content
Text 6	M	L	2.59
Text 5	M	H	2.52
Text 4	H	L	2.19
Text 2	L	L	2.08
Text 7	M	H	2.07
Text 3	H	L	2.0
Text 9	M	H	1.83
Text 8	M	H	1.79
Text 1	L	H	1.56

Relevance to PIE and Second Language Learning

Rewriting texts involves a great deal of cognitive moving parts. Writers bring a certain amount of background knowledge that can affect their decisions in how to paraphrase a text, but the text itself has some influence. By examining the previous texts, how the complexity of the texts plays a role in the successful outcome of paraphrasing tasks may be clearer. The answers to the research questions were helpful in this regard.

1) Do elevated levels in lexical complexity result in low levels of paraphrasing success?

Yes, the copying was more obvious with the lexically complex texts. Texts 3, 4, and 6 were copied much more than text 2, which was low lexical complexity. Also, although the content accuracy for texts 3 and 4 were about the same as text 2, the participants copied a lot more strings to rewrite the texts. Texts 5, 7 - 9, were all medium level lexical complexity, but had mixed results when comparing content accuracy to text 1.

2) Does high grammatical complexity result in low levels of paraphrasing success?

Yes, the grammatically complex text 1 was copied more than text 2, and text 7 was copied considerably more than 6; however, texts 8, 9, and 5 were only copied at about the same

rate as 6, which was supposed to be grammatically less complex. There was also a considerable difference in accuracy between text 1 and 2. Texts 7-9 were also quite low compared to text 6 even though the degree of complexity was only medium to low. Once again text 5 did not fit into the pattern, possibly for the reasons addressed above.

L2 writers do seem to try to make up for gaps in their language when paraphrasing (McDonough et al, 2014). If higher complexity in the text is where the gaps exist, then it is clear that the gaps are filled in with either copied strings or language that is inaccurate in content. Noun-noun phrases, collocations, and noun chunks were specifically targeted for copied strings. The relationships between nouns and noun phrases may represent a more complex knowledge that is not fully understood by L2 writers, which then results in an unsuccessful paraphrasing attempt.

If lexical complexity is an encourager of copying, teachers and educators should be aware of the types of lexical complexity that exist in academic texts, which would be more of an issue for L2 writers, in order to help writers avoid copying long strings of words. The noun chunks and collocations should be discussed more, and how to handle them in paraphrasing can be practiced in the classroom. Just as important to proficient writing is dealing with the grammatical complexity that is too complex for writers to understand or even recognize as higher-level grammar. Often L2 writers seem to be pushed into paraphrasing texts with complex grammar before being exposed to the structures in class. While this may be out of necessity to follow curriculum guidelines, teachers and educators should be aware that extra assistance may be necessary for L2 writers to correctly interpret the high level grammatical structures.

Rewriting texts is not just a vital and frequent activity in academia, it is an activity that brings to bear many of the complex issues and skills required in the writing process. L2 writers

do what is necessary to finish the task as they desire at the time. For some, this may mean copying the text, while others may choose to attempt a reformulation. Based on the findings from this study, it seems that many writers may be more inclined to copy when the text is more complex, especially lexically complex. Texts that were more grammatically complex, but not a problem lexically, seem to encourage some writers to reformulate. These paraphrases were met with mixed results in content accuracy.

References

Anthony, L. (2013). AntWordProfiler (Version 1.4.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Barks, D., & Watts, P. (2001). Textual borrowing strategies for graduate-level ESL writers. In D. Belcher and A. Hirvela (Eds.), *Linking literacies: Perspectives on L2 reading-writing connections*, pp. 246-267. Ann Arbor, MI: University of Michigan Press.

Keck, C. (2006). The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *Journal of Second Language Writing, 15*, 261-278.

McDonough, K., Crawford, W. J., & De Vleeschauwer, J. (2014). Summary writing in a Thai EFL university context. *Journal of Second Language Writing, 24,* 20-32.

Pecorari, D., & Shaw, P. (2012). Types of student intertextuality and faculty attitudes. *Journal of Second Language Writing, 21*, 149-164.

Pennycook, A. (1996). Borrowing others' words: Text, ownership, memory, and plagiarism. *TESOL Quarterly, 30,* 201-230.

Polio, C., & Shi, L. (2012). Perceptions and beliefs about textual appropriation and source use in second language writing. *Journal of Second Language Writing, 21*, 95-101.

Shi, L. (2004). Textual borrowing in second-language writing. *Written Communication, 21*, 171-200.

----- (2012). Rewriting and paraphrasing source texts in second language writing. *Journal of Second Language Writing, 21*, 134-148.

Television. Introducing Issues with Opposing Viewpoints. Bernay, E., edt. Detroit: Greenhaven Press. 2008.

Appendix A

Please rewrite all the texts below using **your own words and sentence structure**. You may use a paper dictionary, but not anything else. Thank You!

1. The responses she gets generally fall into one of two camps. One group says personal responsibility died a slow but certain death in this country a long time ago and that she is just another principled but idealistic dreamer who has yet to accept the inevitability of government censorship.

2. When starting the withdrawal from TV, explain why you are making these changes and that it is not a punishment. The first month will be the most difficult. Children may cry or plead, but you can remain firm if you keep in mind that you are freeing them from an addiction.

3. Another aspect of personal responsibility is taking care of your neighbor's needs. The media's voyeuristic, celebrity-driven entertainment and "news" programming promotes narcissism, not charity.

4. The National Cultural Values Survey reveals a striking correlation between greater exposure to television and permissive moral views. Heavy television viewers (four hours or more per evening) are less committed to character virtues like honesty and charity, and more permissive about sex, abortion and homosexuality.

5. It is also imperative that you help your children learn how to fill the time that they formerly spent watching TV. Work with them to nurture interests, discover hobbies, and explore new possibilities.

6. Americans discuss the merits and harms of television-watching, yet adults consume three to four hours of TV a day. Parents and experts generally agree that excessive television watching is harmful to children, but 70 percent of children have a television in their bedrooms and spend an average of four hours a day watching.

7. The Secret Millionaire, a British series that started broadcasting in October [2006], features a wealthy individual who disguises his or her identity and lives in a working-class area to find a worthy recipient of almost $100,000.

8. Advertisers must diligently defend their right to communicate with American families through the universe of media options the industry helps finance. And they should trust parents to draw sensible lines for their kids.

9. Other issues can intensify the debate. Parents and media-awareness groups have accused the food industry of worsening the childhood obesity problem by specifically targeting children in commercials for junk food. Advertisers insist that parents need to resist buying junk food and turn off the TV if they are unhappy with content. Parents and child-health advocates argue that both advertisers and the broadcast industry must restrict commercials aimed at children-and if they refuse to do so voluntarily, some say, the government must force them to do so.