



6-2010

Textual Appropriation in Engineering Master's Theses: A Preliminary Study

Edward J. Eckel

Western Michigan University, edward.eckel@wmich.edu

Follow this and additional works at: https://scholarworks.wmich.edu/library_pubs



Part of the Library and Information Science Commons

WMU ScholarWorks Citation

Eckel, Edward J., "Textual Appropriation in Engineering Master's Theses: A Preliminary Study" (2010).

University Libraries Faculty & Staff Publications. 1.

https://scholarworks.wmich.edu/library_pubs/1

This Article is brought to you for free and open access by the University Libraries at ScholarWorks at WMU. It has been accepted for inclusion in University Libraries Faculty & Staff Publications by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.



Textual Appropriation in Engineering Master's Theses – A Preliminary Study

Edward J. Eckel
Assistant Professor
University Libraries
Western Michigan University
Kalamazoo, Michigan
edward.eckel@wmich.edu
Telephone: (269) 387-5140
Fax: (269) 387-5180

Abstract

In the thesis literature review, an engineering graduate student is expected to place original research in the context of previous work by other researchers. However, for some students, particularly those for whom English is a second language, the literature review may be a mixture of original writing and verbatim source text appropriated without quotations. Such problematic use of source material leaves students vulnerable to an accusation of plagiarism, which carries severe consequences. Is such textual appropriation common in engineering master's writing? Furthermore, what, if anything, can be concluded when two texts have been found to have textual material in common? Do existing definitions of plagiarism provide a sufficient framework for determining if an instance of copying is transgressive or not? In a preliminary attempt to answer these questions, text strings from a random sample of 100 engineering master's theses from the ProQuest Dissertations and Theses database were searched for appropriated verbatim source text using the Google search engine. The results suggest that textual borrowing may indeed be a common feature of the master's engineering literature review, raising questions about the ability of graduate students to synthesize the literature. The study also illustrates the difficulties of making a determination of plagiarism based on simple textual similarity. A context-specific approach is recommended when dealing with any instance of apparent copying.

Keywords

plagiarism; patchwriting; textual appropriation; information ethics; academic writing;

Introduction

The thesis literature review is where a graduate student is expected to place his or her research results in the context of previous work by other researchers. However, the quality and length of the literature review can vary widely, depending upon the student's ability to both choose relevant literature (Boote & Beile 2005) and write for synthesis (Dong 1998). Hence, for some graduate students, the literature review may be a rich but problematic mixture of original writing and source text appropriated without quotations and accompanied by spotty or nonexistent attribution. Unfortunately for those students, such a patchwork or "cut and paste" way of writing the literature review will often be considered plagiarism when held up to the unstinting scrutiny of a particular institution's policy on research misconduct.

In early 2006, the news media reported that the literature review sections of as many as 55 master's theses at Ohio University's Russ College of Engineering were found to contain copious amounts of verbatim source material (Grose 2006). This brought engineering into the news in a very negative light and resulted in one student losing his master's degree (Linder 2007) and others being required to rewrite their theses. When incidents such as this are publicized, little is usually said about how much material was copied and whether the copying represented the stealing of words or ideas or both (Bouville 2008). An instance of copying, regardless of degree and context, appears to equal a finding of plagiarism in many cases. But despite anecdotal reports like this, how much source text graduate engineering students appropriate and the nature of that appropriation is not well understood.

Scholarly practices of textual appropriation and attribution in academic writing are rarely as straightforward as we usually tell students they are (Parry 1998). Many factors govern how and why scholars and students make use of source texts in their academic writing, including English language facility (Dong 1998; Flowerdew & Li 2007; Yilmaz 2007), cultural background (Pennycook 1996; Shi 2006), facility in basic citation practices (Pecorari 2001; Shi 2006; Shi 2010), and the social requirements of the discipline (Moed 2005). Given the complexity of this potential minefield, it is understandable how both experienced and novice academic writers can make mistakes in how they both use and attribute source texts. It is also understandable how difficult it can be to ascertain whether a given instance of textual appropriation or copying fits an institution's or an individual instructor's definition of plagiarism.

A more nuanced sense of graduate student text use and reuse has the potential to place incidents like Ohio University's in a clearer and more useful context. How well do institutional and disciplinary definitions of textual (i.e. "word for word") plagiarism handle the reality of textual appropriation by graduate students in the writing of background material?

Literature Review

Within the literature on academic writing, two dominant ways of viewing plagiarism or textual appropriation can be discerned. One treats word for word plagiarism regardless of degree in a similar manner as other instances of "academic dishonesty" such as copying on exams or buying papers from online paper mills (Brown 1994; Brown 1996; Carpenter, Harding, Finelli, Montgomery, & Passow 2006; McCabe 1997). Another perspective views it as a continuum of

compositional strategies utilized by both novice and expert writers originating from a complex of reasons, few of which may have anything to do with deception or dishonesty.

Novice writers, including English as a Second Language (ESL) students, may appropriate source material either from an inability to paraphrase (Howard 1999), as a way of adopting the language and vocabulary of a particular discipline through imitation (Barks 2001; Bianchi & Pazzaglia 2007; Clarke 2006; Flowerdew & Li 2007; Howard 1999; Jones & Freeman 2003; McCutchen 1994; Shi 2006), or due to a different cultural perspective on text use (Shi 2006). Howard has termed such behavior “patchwriting” (1999). Pecorari (2003) provides evidence that non-English speaking post-graduate students may engage in problematic text use without intending to deceive.

Even relatively expert writers may engage in textual appropriation for understandable reasons. In a survey conducted by Parkhurst (1990), non-native English speaking scientists report “imitating the style of what they had read” as being essential in their learning to write well in their disciplines. Interestingly, what Parkhurst’s respondents leave unspoken is whether the “style” they are referring to includes generic grammatical constructions or the actual text of their sources. In a letter to the journal *Nature*, the Turkish physicist Yilmaz (2007) unapologetically admits to using verbatim (but attributed) source text without quotes in introductory and background sections to “better introduce the problem” because the introduction needs to “[read] well enough for the paper to be published and read.”

Textual appropriation in academic writing can arise in other ways as well. An increasing body of evidence within linguistics indicates that some formulaic text strings are reused on a common basis in academic writing. Biber, Johansson, Leech, Conrad, & Finegan (1999) call such repeated strings of 3-6 words in length “recurrent lexical bundles” and have tabulated them as making up as much as 21% by frequency of academic writing. While such bundles lack intellectual content per se, they have a distinct functional value as rhetorical framing devices within sentences (Biber et al. 1999; Simpson-Vlach & Ellis 2010). Simpson-Vlach and Ellis (2010) even see such bundles as being useful in helping ESL students incorporate standard academic constructions into their writing.

Is such recurrent language present in the more specific corpus of science and technology writing? Pecorari (2009; 2008) extended the linguistic approach of Biber et al. and Simpson-Vlach and Ellis to biology writing, finding evidence that a number of lexical bundles, including a small number of lengthy ones, can occur in a very limited biology corpus. Sorokina, Gehrke, Warner, and Ginsparg (2006) analyzed academic papers in the online technology collection *arXiv.org*, using a computer program to compare papers for unique matches of seven words or more in length. While their study could not conclude that serious word-for-word plagiarism had occurred in most of the papers with text matches, it demonstrated that repeated or formulaic constructions do occur in technology writing.

In a precursor to the current study, McCullough and Holmberg (2005) used the Google search engine to search strings of text from 210 master’s theses for instances of “potential plagiarism,” which they defined as any suspect phrase that had a match to a document on the web. They

discovered appropriated strings in over 27% of the theses searched, with the highest incidences occurring in the engineering disciplines. While McCullough and Holmberg give no contexts for the copied material, their results suggest that textual appropriation of some type may be a feature of engineering master's writing.

Due to the numerous factors that influence student textual copying besides academic dishonesty, Chandrasoma, Thompson, and Pennycook (2004) have recommended replacing the traditional academic notion of textual plagiarism with the concepts of transgressive or non-transgressive appropriation. Whether a particular instance of copying is right or wrong, they suggest, cannot depend solely upon whether a string of text was copied, but must also come out of the context of the writing. That context includes the language background of the student, the nature of the text appropriated, the nature of the writing assignment, the disciplinary context, and even the instructor's personal viewpoint.

The aim of this study is to gather some preliminary data on patterns of textual appropriation in master's engineering writing. The study was designed to answer the following exploratory questions:

1. How prevalent is the appropriation of verbatim source text in engineering master's theses?
2. Do existing definitions of plagiarism provide a sufficient framework for determining if an instance of copying is transgressive or not?

Methods

The study sample was drawn from the ProQuest Dissertations and Theses Full Text database.

The researcher searched the database for master's theses under the subject heading

“engineering,” limiting the retrieval to English language theses completed within the year 2007.

From the resulting 1961 theses, a random sample of 100 was chosen for study. A sample of this size would have an estimated margin of error of .1933 at a .95 confidence level for the population size (Lenth 2006). While this sample was smaller than would be ideal for maximum statistical rigor, it was large enough to make data collection feasible given time limitations.

As shown in Table 1, the majority of theses were from institutions in the United States. Theses from 47 different institutions were captured in the random sample.

Country	Number of Theses
USA	62
Canada	35
Saudi Arabia	2
India	1
Total:	100

Table 1: 2007 Engineering Master's Theses by Country

Note: The USA total includes 1 thesis from Puerto Rico.

The following methodology was adapted from a procedure developed by the researcher to conduct a plagiarism investigation at his institution and for which the Turnitin service proved ineffective.

Text strings for searching were chosen by visually scanning each page of the literature review or background section and selecting short strings that superficially appeared uncommon or distinctive enough to result in likely hits. Selected text strings ranged from 4 to 7 words in length. Examples of the types of search strings chosen can be seen in Table 2 below. It is important to emphasize that such a method of text selection is ad hoc, even subjective.

Strings were entered into the Google search box as exact phrases surrounded by quotation marks. If too many hits were retrieved, more words from the surrounding text would be added until either a manageable retrieval or a “No results found” message was obtained, the goal being to find one exact match. For practicality, search time was limited to one hour per thesis.

Sample Search Strings
10 by 10 symmetric distance matrix
abstractions of real or envisioned situations
calculating the flexibility of mitred
dispersed multiphase-multicomponent interactions
embed the extracted quality features
facilitated deterioration due to resin
gain starts from an unsaturated value
HandelC syntax is based on conventional
ignore spatial variability of land uses
leadership has been perceived as a soft skill

Table 2: Sample Search Strings Used in Google

Basic bibliographic information (title, author, year of publication or posting, and URL) was recorded for each online source match. Each source was visually scanned around the matching text to see how much surrounding text from the source also matched text from the thesis. All lexical and function words were counted word-for-word, excluding chemical and mathematical formulae and numbers. For convenience, strings longer than a paragraph in length were pasted into Microsoft Word and counted using Word’s “Word Count” function.

It was impossible to be certain whether a source found using Google was the actual original source of the search string. In many cases, there may very well have been an Ur-text that predated the source found online. For the purposes of the study, a text match was assumed to be the source of the search string if:

- It was cited by the thesis author, or
- No other sources for the text string were found, and
- It was published or posted before or during 2007 (because some thesis authors did reference documents of 2007 vintage).

Since this study focuses on text appropriated from other authors rather than so-called “self-plagiarism,” text matches to other works by a thesis author, such as articles or publically posted copies of a thesis, were excluded from this study. Also excluded were short text strings (e.g. 5-8 words) that matched source text but not the subject/disciplinary context of the source. Without a match to the subject context, it would be impossible to verify whether the text match came about through deliberate copying or from the normal recycling of text and speech that comes from immersion in English-speaking culture. A match within the same subject or disciplinary context would at least provide tentative evidence for appropriation. Finally, copied background figures or charts from reference books were excluded because such appropriation was beyond the scope of this study.

Since attribution or the lack thereof is relevant to a finding of plagiarism, the presence of attribution for the copied textual material was also recorded. This would at least give some sense of student skill levels with attribution.

Results

Table 3 shows the results of the text searches organized into four tentative degrees of appropriation, based upon the number of exact text matches found and the length of the longest text match per thesis. Theses for which zero to one text match was found were classified at the low end of textual appropriation for comparison. It must be acknowledged that some appropriated text would not be discovered due to the ad hoc nature of the initial search string selection and because potential source materials might not have been indexed by Google. Hence, the actual number of theses with zero to one text match could be smaller.

Text matches of one page or more in length were categorized at the other extreme. In between these maximum and minimum categories were two intermediate categories. It is important to note that those who appropriated more strings also copied longer strings.

Degree of Textual Appropriation	Number of Theses	Longest String (in words)	
		Max	Min
0 to 1 string	47	29	0
2-4 strings (totaling < 1 page)	19	119	9
5 or more strings (totaling < 1 page)	27	224	17
One page or more	7	2447	271
Total:	100		

Table 3: Results by Degree of Textual Appropriation

The 28 theses with the longest text matches (those over 50 consecutive words in a row or about a fifth of a double-spaced page) are shown in Table 4. Each thesis is listed by the random number used to generate the sample. The length of the longest verbatim string is listed, whether attribution was provided for that string, along with the total number of verbatim strings found for that author.

Random No.	Max String Length (Words)	Attribution	Total No. Verbatim Strings
0556	2447	yes	3
5277	671	Yes - not in context	3
2966	641	yes	6
1640	518	Citations copied from source	8
1374	384	Yes - not in context	4
3527	312	yes	9
0720	271	yes	7
0785	224	yes	5
4323	199	yes	11
3296	122	yes	5
1956	120	no	6
5314	119	yes	2
5438	111	yes	7
3509	110	no	5
4557	104	no	10
1286	96	yes	6
1650	91	yes	13
0532	91	yes	8
3580	82	yes	6
5462	79	yes	6
1180	76	no	7
2011	68	yes	9
3929	68	no	5
4515	61	yes	5
1302	56	no	13
1009	54	no	7
4542	54	yes	4
0628	50	Yes - not in context	8

Table 4: Verbatim String Data (Over 50 Consecutive Words)

The longest string for each of the engineering master's students (excluding the very longest string, to give a clearer picture of the dataset) is shown in Figure 1. Examples of the types of text matches seen in this study, with their sources, are shown in Figures 2 and 3 and Table 5.

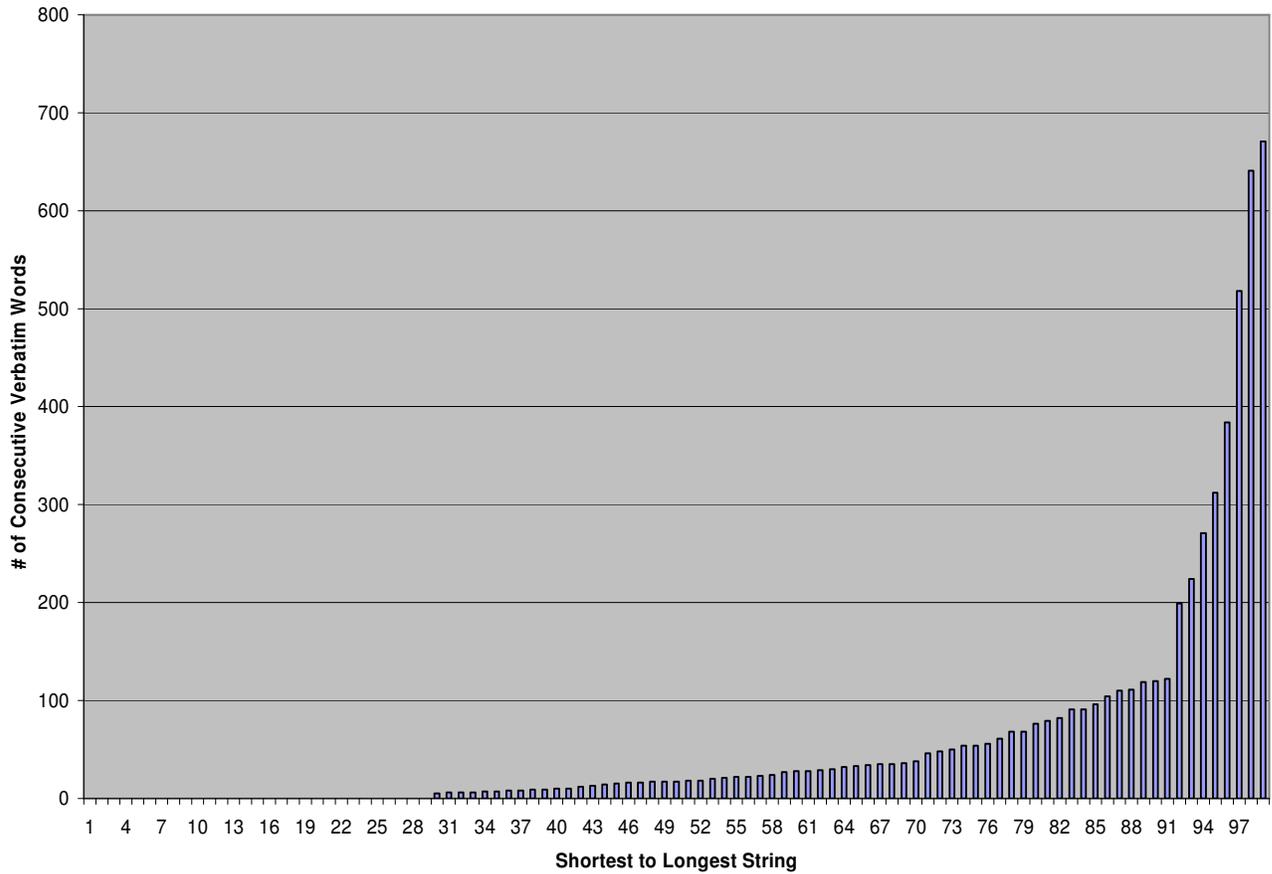


Figure 1: Distribution of Longest Text Matches (Per Thesis)
(Very longest string excluded)

2.1

Purely elastic medium

The general solution for the displacements and stresses given by (complex) combinations

$$u \equiv u_x + iu_y, u_z,$$

$$\Gamma_1 \equiv \Gamma_{xx} + \Gamma_{yy}, \Gamma_2 \equiv \Gamma_{xx} - \Gamma_{yy} + 2i\Gamma_{xy}, \Gamma_{zz}, S_z \equiv \Gamma_{zx} + i\Gamma_{yz} \quad (1)$$

for a purely elastic transversely isotropic medium can be given in terms of harmonic potential functions $F_j \equiv F(x, y, z_j)$, $j = 1, 2, 3$, where $z_j = z/c_j$ in the form, [19]:

$$u = K \left(\sum_{j=1}^3 F_j + iF_3 \right), \quad u_z = \sum_{j=1}^3 m_j \frac{\partial F_j}{\partial z},$$

$$\Gamma_1 = 2 \frac{\partial^2}{\partial z^2} \sum_{j=1}^2 (C_{66} c_j^2 - a_j) F_j, \quad \Gamma_2 = 2C_{66} K^2 \left(\sum_{j=1}^2 F_j + iF_3 \right), \quad (2)$$

$$\Gamma_{zz} = \frac{\partial^2}{\partial z^2} \sum_{j=1}^2 a_j c_j^2 F_j = -D \sum_{j=1}^2 a_j F_j, \quad S_z = K \frac{\partial}{\partial z} \left(\sum_{j=1}^2 a_j F_j + iC_{44} F_3 \right),$$

Figure 2: Example of Verbatim Mathematical Description

Source: Karapetian, E. et al. 2002. The principle of correspondence between elastic and piezoelectric problems.

Archive of Applied Mechanics 72: 564 – 587

Reprinted with permission of Springer Science+Business Media

II. Controller Development

A. Problem Statement

We consider the system of the form

$$\dot{x}_1 = f_1(x) \quad (7)$$

$$\dot{x}_2 = f_2(x) + B_2(x)u \quad (8)$$

where $x_1 : R \rightarrow R^{n_1}$ and $x_2 : R \rightarrow R^{n_2}$, $n_1 + n_2 = n$, and $x(t) = [x_1^T(t) x_2^T(t)]^T \in X \subset R^n$ is the state vector. The $u(t) \in U \subset R^m$

Figure 3: Example of Verbatim Theoretical Description

Source: Lu, Ping. 1994. Nonlinear predictive controllers for continuous systems.

Journal of Guidance, Control, and Dynamics 17(3): 554.

Reprinted with permission of the American Institute of Aeronautics and Astronautics

String Type	Sample Text Match	Source
Bullet Points	<p>“The recent trends in improving the impedance bandwidth of microstrip antennas can be broadly divided into the following categories:</p> <p>(i) Various geometries and perturbations to introduce multiple resonances as well as input impedance matching,</p> <p>(ii) Genetic Algorithm (GA) based optimization of antenna geometries,</p> <p>(iii) Photonic Band Gap (PBG) structures used as printed antenna substrates,</p> <p>(iv) Frequency Selective Surfaces (FSS) used as multilayered substrate or ground Plane.”</p>	<p>Guha, Debatosh. 2003. Broadband design of microstrip antennas: Recent trends and developments. Invited paper. <i>Mechanics, Automatic Control and Robotics</i>. 3(15): 1084.</p>
Common Knowledge	<p>“...the job of the supply chain management system is to find a way to fill current and anticipated orders at lowest cost while meeting customer service constraints. Pricing and Revenue Management assumes that variable costs and capacity availabilities are fixed, and it looks to find the set of prices and customer allocations that maximizes profitability, subject to these constraints.” – from p.13</p>	<p>Phillips, Robert Lewis. 2005. <i>Pricing and revenue optimization</i>. Stanford, Calif. : Stanford Business Books: 13.</p>
Definition	<p>“The cured glass transition temperature (T_g) of a <i>polymeric material</i> [changed to “polymer” by thesis author] is the temperature at which it changes from a rigid glassy solid into a softer, semi-flexible material.”</p>	<p>Campbell, Flake C. 2004. <i>Manufacturing processes for advanced composites</i>. Oxford: Elsevier: 96.</p>
Introductory Text	<p>“The following ground rules and assumptions are defined so that a knowledgeable person can reproduce the results presented in this study. They identify constraints placed on the process allowing an accurate estimate of Complex GA Aircraft reliability, which is the primary purpose of this study. They also define failure, isolate factors from the analysis that may obscure hardware failure, and aid in simplifying the analysis. Although no two analysts will perform an analysis the same way, it is believed that the basic ground rules and assumptions used would not grossly deviate from those presented here. In this analysis, failure occurs when the inherent ability of a component to perform its intended function is lost and therefore could lead to a loss of an aircraft's system/subsystem function.”</p>	<p>Turnbull, Andrew, & Pettit, Duane. 2001. <i>General Aviation Aircraft Reliability Study: Section F. Data Collection</i>. Contractor Report No. NASA/CR-2001-210647. Hampton, VA, National Aeronautics and Space Administration. http://www.tpub.com/content/nasa2001/NASA-2001-cr210647/</p>
List	<p>“... sand, silt, clays, Giardia lamblia and Cryptosporidium cysts, algae, and some bacterial species.”</p>	<p>Lahlou, Mohamed. 1999. <i>Membrane Filtration: Tech Brief: A National Drinking Water Clearinghouse Fact Sheet</i>. Morgantown, WV: National Drinking Water Clearinghouse: 2-3. On the web at http://www.nesc.wvu.edu/pdf/DW/publications/ontap/tech_brief/TB12_Membrane.pdf</p>

Table 5: Sample Text Matches, with Sources

As the Google searches were conducted, it became apparent that a number of verbatim strings found in the theses were appropriated across multiple websites (Table 6). Following the research methodology, the search strings were usually lengthened when this happened to narrow down to a specific source text. However, the presence of textual appropriation across the Web seemed important enough to record, as it suggested that some of the textual appropriation seen here may be part of a larger pattern of text reuse in engineering writing. Table 6 gives the longest twenty

strings with the number of web hits from Google. As can be seen in that table, some fairly lengthy strings could have several hundred hits, including one string 17 words in length that had 353 web hits.

Search String	No. of Words	Hits
skin effect causes the effective resistance of the conductor to increase with the frequency of the current	17	353
girth of a graph is the length of the shortest cycle in the graph	14	8
girth of a graph is the length of the shortest	10	572
real-valued function whose value depends only on the distance	10	9
proportional to the phase difference between the clock and data	10	4
can be used to simulate human-induced climate change	9	81
mobile network operators continuously seek new and innovative ways	9	7
magnetic field is normal to the plane of incidence	9	7
integration of mechanical elements, sensors, actuators and electronics	8	568
graph G consists of a set of vertices	8	527
equal number of positively and negatively charged particles	8	286
objective is to find a feedback control law	8	70
where w is the vector of conservative variables	8	47
planning, design, construction, and maintenance of transportation	7	1,750
fly ash, ground granulated blast-furnace slag	7	1,100
greatest range of motion of any joint	7	578
phase difference between the clock and data	7	511
operators continuously seek new and innovative ways	7	190
objective is to find a feedback control	7	110
Transistor is a three-terminal semiconductor device	7	105

Table 6: Strings with Multiple Hits in Google (longest twenty)

Discussion

Many standard or institutional definitions of plagiarism give the presence or absence of attribution as a key condition in any finding of plagiarism, including the Modern Language Association (2009), National Science Foundation (2002), and Western Michigan University (2006) where this researcher is located. While source text has pretty clearly been appropriated in the theses examined here (Tables 3 and 4), the context of the appropriation is very unclear.

Many of the authors in Table 4 provided some type of attribution for the copied text. How can one categorize such attributed copied material? Some authors appear to have imported the original citations along with the verbatim text from their sources (e.g. author 1640). Others provided citations that were not contiguous, but rather in a preceding or succeeding paragraph, on a separate page, or only in the bibliography. Finally, some writers in this study cited nothing for the text they copied. Little can be concluded from such a variety of attribution patterns except to concur with Pecorari's point (2001) that the common advice to students to avoid plagiarism by simply "providing citations" is an ineffectual one.

Some definitions of plagiarism, such as the one by the Australian Vice-Chancellor's Committee (Australian Vice-Chancellors' Committee, 1997) require determining whether or not the copying was intentional or accidental. Such a definition was not useful in this context as it was impossible to ascertain the authors' intentions. Also, individuals, even within the same discipline, might disagree about whether a particular instance of copied text would count as plagiarism, and if so, whether it was a serious or relatively minor offense (Chandrasoma et al. 2004; Crocker & Shaw 2002; Flint, Clegg, & Macdonald 2006; Shi 2006).

The length of the copied strings might provide a comparative sense of transgression. However, published metrics based upon word count vary from three (Chekola & Shaw 2007; Drum 1986; Hodges, Whitten, & Connolly 1962), four (Hexham 2005), five (Roig 1999), and seven (Sorokina et al. 2006) words in a row. Such guidelines are not only arbitrary, as Bouville (2008) argues, they also do not take into account repeated language which research has shown is a common characteristic of academic writing (Biber et al. 1999; Pecorari 2009; Pecorari 2008;

Simpson-Vlach & Ellis, 2010). In addition, word count metrics can not be applied to appropriated mathematical formulae such as those in Figures 2 and 3 with few lexical words to count.

To make a determination of patchwriting, one would need to do more than simply demonstrate the apparent copying of isolated sentences or text strings. Patchwriting is known predominantly as a strategy of novice or ESL writers. In this study, both the language background and writing experience of the master's students was indeterminable. Also, because this study focused only upon exact text matches, some incompletely paraphrased material which would also indicate patchwriting may have been missed in the Google searches. Given the study limitations, trying to determine if appropriated sections were patchwritten would have been guesswork.

None of the lengthy verbatim strings referenced in Table 4 fit the description of repeated language or lexical bundles described by Biber et al. (1999), Pecorari (2009; 2008), and Simpson-Vlach and Ellis (2010) since such recurrent language is generally acknowledged to consist of relatively short text strings. It is possible that the shorter text matches not analyzed in this study might have resulted from common constructions, and therefore be non-transgressive in nature. Furthermore, the examples of text strings in Table 6 that are appropriated in multiple places on the Web do suggest that some fairly lengthy text strings within technical writing may be repeated language. Without the benefit of additional data on repeated language within engineering, however, neither of these two possibilities could be verified.

Perhaps the most important criterion for a finding of transgressive plagiarism would be whether ideas versus words were copied in the appropriated strings? While some researchers argue that the background nature of literature reviews precludes original thought and thereby the stealing of ideas (Bouville 2008), a claim can be made that the purpose of the literature review is to make an intellectual argument that contextualizes the current research within previous work. That said, differentiating between ideas and words in a text match is not a clear-cut procedure. As can be seen in Table 5, textual material such as the “Definition” and “List” strings are clearly background material with little intellectual content. The “Introductory Text” string in that table is a little more ambiguous. To attempt to discern the copying of ideas versus words would have required a method of coding each text string as unique idea, common knowledge or background content. Such a coding scheme would be difficult without disciplinary knowledge and expertise this researcher lacked, especially working from a random sample of theses covering such a broad range of engineering disciplines. Also, what counts as “common knowledge” can vary even between individuals within the same discipline (Shi 2010). Indeed, it is not clear that any such distinctions could be reliably made by anyone outside his or her own specific sub-discipline.

Conclusion

The results of this study must be interpreted with caution. The small sample of theses examined, the ad hoc method of selecting search strings and the focus on exact text matches combine to make any generalization extremely tentative. Furthermore, the choice to analyze only the longest text matches at the exclusion of the shorter copied strings may give a simplified view of textual appropriation in these theses. Certainly, a standardized method of selecting initial search

strings and a more in-depth analysis of all text matches could give a much better idea of the ways in which graduate engineering students use source text. Such an improved study would help inform future discussions of the ethics of textual appropriation within engineering and the nature of the academic writing process.

Initially, the researcher assumed that identifying text matches using Google would be enough to make some conclusions regarding the prevalence of transgressive plagiarism in engineering master's theses. This was not borne out by the present study. In practice, existing conceptions of plagiarism can actually be quite problematic to apply out of context given the number of factors that govern the use of source texts by graduate students (Chandrasoma et al. 2004; Crocker & Shaw 2002; Pecorari 2003; Pecorari 2001; Shi 2010). The value of this study is in highlighting the difficulties inherent in making a determination of plagiarism based upon a simple text match between documents.

While the results suggest that textual borrowing may be a common feature of the master's engineering literature review, the nature of the borrowing could not be determined due to limitations in the methodology. That said, the more extreme textual appropriation seen here appears very questionable. It raises questions about engineering graduate students' abilities to synthesize the literature. Pecorari (2008) has shown that even in a limited disciplinary corpus like biology there are alternatives for rephrasing material. While a graduate student would be hard pressed to justify such extreme copying, it is understandable within the context. Many graduate students are still learning how to contextualize their research without an overreliance upon the exact language of their sources. Obviously, some will do this better than others. For

some, the literature review will remain something done to fulfill a requirement, a mere listing of what others have done “with no attempt to shovel the glimpse into the ditch of what each one means” (Dylan 1965).

In all writing, the requirement of originality exists in tension with the competing need to utilize a common language to which many writers and scholars have contributed textual material. Such tension impinges perhaps more fully upon novice or ESL writers. A context-specific approach that takes factors such as those discussed above into account is needed when dealing with any instance of apparent copying.

Acknowledgments

The author gratefully acknowledges the assistance of several colleagues in the preparation of this manuscript. An early draft of this manuscript was presented at the ACRL Science and Technology Section Research Forum at the 2009 American Library Association Annual Conference. At that program, Julia Gelfand provided pointed and astute commentary on my paper that prompted further substantive revisions. Carrie Leatherman, natural science librarian at Dwight B. Waldo Library and Dr. Lewis Pyenson, Dean of the Graduate College at Western Michigan University read and gave supportive feedback on the paper before that presentation. Thanks to the anonymous reviewers for their assistance in fine-tuning this article.

References

- Australian Vice-Chancellors' Committee. (1997). Joint NHMRC / AVCC statement and guidelines on research practice. <http://www.nhmrc.gov.au/grants/policy/researchprac.htm>. Accessed 20 January 2010.
- Barks, D. (2001). Textual borrowing strategies for graduate-level ESL writers. In D. Belcher, & A. Hirvela (Eds.), *Linking literacies: Perspectives on L2 reading-writing connections* (pp. 246-267). Ann Arbor: University of Michigan Press.
- Bianchi, F., & Pazzaglia, R. (2007). Student writing of research articles in a foreign language: Metacognition and corpora. In R. Facchinetti (Ed.), *Corpus linguistics 25 years on* (pp. 259-287). Amsterdam and New York: Rodopi.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, England: Pearson Education Limited.
- Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher*, 34(6), 3-15.
- Bouville, M. (2008). Plagiarism: Words and ideas. *Science and Engineering Ethics*, 14(3), 311-322.
- Brown, B. S. (1994). The academic ethics of graduate engineering students. *Chemical Engineering Education*, 28, 242-243, 265.
- Brown, B. S. (1996). A comparison of the academic ethics of graduate business, education and engineering students. *College Student Journal*, 30(3), 294-301.

- Carpenter, D. D., Harding, T. S., Finelli, C. J., Montgomery, S. M., & Passow, H. J. (2006). Engineering students' perceptions of and attitudes towards cheating. *Journal of Engineering Education, 95*(3), 181-194.
- Chandrasoma, R., Thompson, C., & Pennycook, A. (2004). Beyond plagiarism: Transgressive and nontransgressive intertextuality. *Journal of Language, Identity, and Education, 3*(3), 171-193.
- Chekola, M., & Shaw, S. (2007). Plagiarism: What it is and how to avoid it. http://www.mnstate.edu/gracyk/expectations_of_students/plagiarism_warning.htm. 14 April 2010.
- Clarke, R. (2006). Plagiarism by academics: More complex than it seems. *Journal of the Association for Information Systems, 7*(2), 91-120.
- Crocker, J., & Shaw, P. (2002). Research student and supervisor evaluation of intertextuality practices. *Hermes, 28*, 39-58.
- Dong, Y. R. (1998). Non-native graduate students' thesis/dissertation writing in science: Self-reports by students and their advisors from two U.S. institutions. *English for Specific Purposes, 17*(4), 369-390.
- Drum, A. (1986). Responding to plagiarism. *College Composition and Communication, 37*(2), 241-243.
- Dylan, B. (1965). Gates of Eden. In *Bringing It All Back Home*. Audio recording. New York: Columbia Records.

Flint, A., Clegg, S., & Macdonald, R. (2006). Exploring staff perceptions of student plagiarism.

Journal of Further and Higher Education, 30(2), 145-156.

Flowerdew, J., & Li, Y. (2007). Language re-use among Chinese apprentice scientists writing for publication. *Applied Linguistics*, 28(3), 440-465.

Grose, T. K. (2006, November). The burden of plagiarism. *ASEE Prism*, 16, 36-39.

Hexham, I. (2005). Academic plagiarism defined.

<http://www.ucalgary.ca/~hexham/study/plag.html>. 12 November 2009.

Hodges, J. C., Whitten, M. E., & Connolly, F. X. (1962). *Harbrace college handbook*. New York: Harcourt, Brace & World Inc.

Howard, R. M. (1999). *Standing in the shadow of giants: Plagiarists, authors, collaborators*. Stamford, CT: Ablex Pub.

Jones, A. A., & Freeman, T. E. (2003). Imitation, copying, and the use of models: Report writing in an introductory physics course. *IEEE Transactions on Professional Communication*, 46(3), 168-184.

Lenth, R. V. (2006). Java applets for power and sample size. Computer software.

<http://www.stat.uiowa.edu/~rlenth/Power>. 4 April 2008.

Linder, S. (2007, March 28). Ohio University revokes master's degree: Student found guilty of plagiarism. *Ohio University Outlook*. <http://www.ohio.edu/outlook/06-07/March/457n-067.cfm>. Accessed 3 August 2009.

McCabe, D. L. (1997). Classroom cheating among natural science and engineering majors.

Science and Engineering Ethics, 3(4), 433-445.

McCullough, M., & Holmberg, M. (2005). Using the Google search engine to detect word-for-word plagiarism in master's theses: A preliminary study. *College Student Journal*, 39(3), 435-441.

McCutchen, C. W. (1994). Plagiarism: A tale of telltale words. *Journal of Information Ethics*, 3, 48-50.

Modern Language Association. (2009). *MLA handbook for writers of research papers* (7th ed.).

New York: Modern Language Association of America.

Moed, H. F. (2005). *Citation analysis in research evaluation*. Dordrecht, The Netherlands:

Springer.

National Science Foundation. (2002). *Research Misconduct Regulation*. 45 CFR 689.

<http://www.nsf.gov/oig/resmisreg.pdf>. Accessed 11 December 2009.

Parkhurst, C. (1990). The composition process of science writers. *English for Specific Purposes*,

9(2), 169-179.

Parry, S. (1998). Disciplinary discourse in doctoral theses. *Higher Education*, 36(3), 273-299.

Pecorari, D. (2001). Plagiarism and international students: How the English-speaking university

responds. In D. Belcher, & A. Hirvela (Eds.), *Linking literacies: Perspectives on L2*

reading-writing connections (pp. 229-245). Ann Arbor: University of Michigan Press.

- Pecorari, D. (2003). Good and original: Plagiarism and patchwriting in academic second-language writing. *Journal of Second Language Writing, 12*(4), 317-345.
- Pecorari, D. (2008). Repeated language in academic discourse: The case of biology background statements. *Nordic Journal of English Studies, 7*(3), 9-33.
- Pecorari, D. (2009). Formulaic language in biology: A topic-specific investigation. In M. Charles, D. Pecorari & S. Hunstan (Eds.), *Academic writing: At the interface of corpus and discourse* (pp. 91-106). London ; New York: Continuum.
- Pennycook, A. (1996). Borrowing others' words: Text, ownership, memory, and plagiarism. *TESOL Quarterly, 30*(2), 201-230.
- Roig, M. (1999). When college students' attempts at paraphrasing become instances of potential plagiarism. *Psychological Reports, 84*(3 PART 1), 973-982.
- Shi, L. (2006). Cultural backgrounds and textual appropriation. *Language Awareness, 15*(4), 264-282.
- Shi, L. (2010). Textual appropriation and citing behaviors of university undergraduates. *Applied Linguistics, 31*(1), 1-24.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research [advance access publication April 14, 2010]. *Applied Linguistics, 1*-26.
- Sorokina, D., Gehrke, J., Warner, S., & Ginsparg, P. (2006). Plagiarism detection in arXiv. *Sixth International Conference on Data Mining, 2006. ICDM '06*. Hong Kong. 1070-1075.

Western Michigan University. (2006). *Research misconduct policy and procedures*

Yilmaz, I. (2007). Plagiarism? No, we're just borrowing better English. *Nature*, 449(7163), 658.