

Opinion article: Plagiarism in three acts

Associate Professor Eric Duff Wrobbel
Southern Illinois University, Edwardsville, USA
ewrobbe@siue.edu

Keywords:

Plagiarism, plagiarism definitions, policy

Abstract

Three brief studies are conducted regarding plagiarism in the age of the internet. The first identifies the disparity between cited and un-cited examples of brief media catchphrases used on the internet. The second uses a thesaurus to generate alternative wordings of established definitions and tests them for legibility. The third asks respondents to spontaneously generate definitions for common concepts and uses Google to test those spontaneous definitions for plagiarism. In all three cases, results are such that they call into question current views on the nature of plagiarism itself.

Introduction

First, allow me to put all my cards on the table. I'm from Southern Illinois University. Yes, you remember correctly – we were the ones in the news over purported plagiarism problems among our senior administrators, and then again for, of all things, plagiarising the definition of plagiarism in a draft of our new plagiarism policy (Weinhold, 2009). Great stuff for late-night comedy had it not been so painful. Here is my problem. Drafting a new plagiarism policy was a response to the earlier problems. I knew the people on the committee. They volunteered. The job was challenging, time-consuming, and thankless, but they were willing to take it on because they were people of integrity and they believed they could make things right. As such, the suggestion that they would have plagiarised the definition of plagiarism is not so much humorous as it is unimaginable and absurd. Clearly, there must be another explanation. I will confess that I immediately 'Googled' the definition and it did, indeed, seem plagiarised. While not identical to any others, it was unnervingly close to one written by another university and (properly) re-used on literally dozens of other websites, so I will admit that things looked bad.

It is estimated that college-educated speakers have a vocabulary of approximately 17,000 words (Zechmeister, Chronis, Cull, D'Anna & Healy, 1995). The plagiarism definition that was reportedly plagiarised was ten words long. Following a mathematical understanding of plagiarism (Falkofske, 2006), the random chance probability for duplicating something of this length should be effectively zero. The math looks like this:

$$(1/17000) \times (1/17000) \times (1/17000) \times (1/17000) \times (1/17000) \times (1/17000) \times (1/17000) \times (1/17000) \times (1/17000) \times (1/17000) = 4.96033 \times 10^{-43}.$$

This is a decimal point, followed by 43 zeros, and then a 496 – a number so ludicrously small as to be essentially zero, so what was I to think? Easy – I thought the math was wrong, or perhaps more accurately, that it did not apply. My academic background is in communication, and specifically in the ethnomethodological tradition of conversation analysis, so instead of considering the math, I thought instead about the *language*. The question I asked myself was simply this: how else could they have said it? When I asked this question, I found that any answer I gave that was clear and coherent sounded just like the one's my colleagues had produced. I came up with almost the same vocabulary, and used almost the same syntax. Every definition I produced that was different enough to be clearly my own was also complex, convoluted, stilted, and even silly-sounding. It was at this point that I undertook the following three experiments.

Experiment #1 – The catchphrase

One thought that occurred to me was that we might be confounding plagiarism and ubiquity. In other words, are there phrases that we've just heard so often that they become part of our vocabulary? How would we test for this? What I did no doubt provided rather imperfect evidence, but it is certainly suggestive. Here is what it was. I identified a number of short phrases from the media – mostly from movies and advertising. I got these by inputting 'popular catchphrases' into the search engine Google, then randomly selecting 15 of the more frequently noted ones. I put each catch phrase I chose into Google within quotation marks. This generated a list of cites containing the exact phrase. I recorded that number. I then ran the same search a second time. On the second search, I also included the source of the phrase and used Boolean search operators to limit the search to only those sites that included *both* items. My presumption was that any site including both has most likely cited the source (a person, film, television advertisement etc.) for the phrase, while those without both have most likely not cited the source. Put differently, those sites without both have plagiarised the phrase. What follows is a list of the phrases, followed by the original source (or in one case, three sources), then the number of properly cited pages over the number of un-cited pages, and finally, the percentage that were properly cited and thus not plagiarised.

- "You complete me" + "Jerry Maguire" 9610/467000 (2%)
- "Where's The Beef?" + "Wendy's" 1740000 + "Walter Mondale" 1030 + "Clara Peller" 296000 total 2037030/4920000 (41%)
- "Hasta la vista baby" + "Terminator 2" 16100/690000 (2%)
- "Wax on, wax off" + "Karate Kid" 34700/212000 (16%)
- "Less Filling – Tastes Great" + "Miller Lite" 1900000/1680 (0.0009%)
- "You can't handle the truth" + "A Few Good Men" 226000/42900000 (0.5%)
- "A little dab'll do ya" + "Brylcreem" 662000/11200 (2%)
- "You had me at hello" + "Jerry Maguire" 10300/153000 (7%)
- "Show me the money" + "Jerry Maguire" 82400/1110000 (7%)
- "Go ahead, make my day" + "Dirty Harry" 11800/149000 (8%)
- "Houston, we have a problem" + "Apollo 13" 18800/255000 (7%)
- "May the force be with you" + "Star Wars" 185000/697000 (26%)
- "A mind is a terrible thing to waste" + "United Negro College Fund" 119000/61400 (52%)
- "Frankly my dear, I don't give a damn" + "Gone with the wind" 58900/266000 (22%)
- "Melts in your mouth and not in your hand" + "M & M" 292/10900 (3%)

What is clear from the (admittedly very non-scientific) data above is that the internet is full of plagiarised material – probably not a surprising finding to anyone. On average,

only 1 in 8 times were the above catchphrases cited properly, with the best case being cited about half the time (52%), and the worst case cited much less than one time in a thousand (0.0009%). From this, we might conclude either that plagiarism is a far bigger problem than we'd imagined, or that something else is going on. I decided to try another experiment.

Experiment #2 – The dictionary

For my second experiment, I asked the question, how else could they have said it? To address this, I first went to *Dictionary.Com* and looked up the word *plagiarism*, which is defined as “the unauthorized use or close imitation of the language and thoughts of another author and the representation of them as one's own original work.” I used the capitalisation rules found in the current edition of the APA manual to identify which words were considered unimportant, then went to *Thesaurus.Com* and collected every synonym for every important word in the definition. For example, *Thesaurus.Com* lists all of the following synonymous with the first important word in the definition of plagiarism, which is *unauthorized*: crooked, dirty, illegal, illegitimate, no-no, off base, out of bounds, out of line, over the line, pirated, shady, unapproved, unconstitutional, under the table, unjustified, unlawful, unofficial, unsanctioned, unwarranted, wildcat, and wrongful. If these words are all synonyms, I presumed that they would be more or less interchangeable, so I selected the first synonym listed for each important word (non-important words were not changed, and appear in bold) in the original definition and put them together in the same order as their original counterparts.

The result is ***the crooked accepting or abutting Xerox of the announcement and aim of another biographer and the apperception of them as a person's endemic avant garde act.*** Not very clear, is it? I tried again with the second synonym listed for each word, and got ***the dirty adapting or across the street apery of the argument and anxiety of an additional columnist and the appreciation of them as all her breaking new ground application.*** Results with the third word from each list were no better, nor the fourth. All such attempts produced gibberish.

As my final (somewhat desperate) attempt to make this experiment produce a functional result, I allowed myself the luxury of selecting what was, in my opinion, the most appropriate word. Surely, with free reign, I could produce an intelligible definition. Here is my best result: ***the unsanctioned appropriation or proximate mimicry of the phraseology and views of some other writer and the representation of them as a person's personal creative product.*** Still not very good, is it?

So all this begs a question. If the goal in generating a definition is to pick the best words, the clearest words, the most universally understood words, then just how many words can one really choose from? If it is important that I clearly express the concept of “unauthorised,” then how many words can I really choose from and still do a credible job? Does “shady” really convey the same meaning? Do most people know what “unsanctioned” means? In other words, despite the fact that English is the most synonym-laden language there is, some words are just clearly better than others, and most people know which words those are.

Experiment #3 – The survey

For my third experiment, I actually did an experiment, complete with IRB approvals and all the trimmings. I contacted the presidents/chairs of the faculty senates from nine state universities in my home state and solicited their participation. Four agreed. These four were sent a packet of surveys to administer to the members of their

respective senates. Along with collecting a bit of demographic information and dispensing with the requisite informed consent forms, participants were told that I was interested in discovering the degree to which faculty members across various institutions and disciplines shared common understandings of a number of critical cross-curricular concepts. Each was asked to write a single-sentence definition of the word or concept provided to them by the study administrator, and to make the definition as clear and straightforward as possible to a general audience – something they might include on a syllabus, a policy sheet, or a website. They were also asked to limit the sentence to no more than 30 words. All were asked to define the same concept – plagiarism. Those administering the survey did so at the beginning of a regularly occurring meeting of their respective senates, and did so unannounced.

Three of the four institutions actually complied, and as a result, I ultimately collected 44 spontaneously-generated definitions of plagiarism, all written by people with advanced degrees, most probably with at least modest experience in policy construction, and all most certainly having a working familiarity with the concept of plagiarism. Some of them seemed a bit idiosyncratic (*As college students, you are expected to be honest in whatever you say or write*) but most are what one would expect (*when an individual takes another person's thoughts, words, or ideas for their own without giving proper credit*). All 44 of these were input into a basic word-processing program, and then I made a series of comparisons between them. What follows are some of my findings.

First, when added together, the 44 definitions contained a total of 720 words. However, most of these words were duplicated. The total number of *unique* words is 181, which gives a ratio of *unique* words to words of 1:4. While some words (26) were only duplicated once (including *product, stating, unacknowledged, and misrepresentation*), there were 17 words (including *using, work, own, and another's*) that were repeated 11 or more times. Many of these would be expected, such as *and* appearing 15 times, and *is* appearing 13 times. More interesting, however, is that both *using* and *work* each appeared 25 times in 44 definitions. In other words, with 44 different people working spontaneously, independently, and without any access to outside materials, there was a remarkably high chance that participants would choose many of the exact same words. There were also several multi-word clusters that appeared more than once. The expressions “the use of another’s work” and “someone else’s words” both appeared identically twice, and “someone else’s work” was duplicated exactly in 4 of the 44 definitions. “As your own” appeared 8 times, and “as their own” appeared 2 more times.

Here is, I suspect, the most interesting finding. One of the hand-written definitions I received was as follows: *representing someone else's words or ideas as your own*. Remember – this was generated for me spontaneously by someone wholly unprepared. If you put this nine word definition into any free internet-based plagiarism checker, you will discover, as I did, that this *exact* definition appears on well over 100 websites, roughly 2/3 of which are from academic institutions. As best I can determine, this definition was first written by someone named Edlund (n. d.), who says it was derived from one at Cal State LA, and it was first put into a policy document by Stinchcomb (n. d.). Only one of the over 100 websites mentions any of these, however. In other words, over 60 universities and one faculty senator working off the top of his/her/their head(s) either plagiarised this definition, or something else is going on.

What else is going on?

As a student, not of literature, but of conversation, here is what I wonder. As the lines between literature and mass media continue to blur, as texting more and more replaces talking, as sound bytes and slogans replace spontaneity, might we need to reconsider the nature of plagiarism? After all, the good folks at Verizon *want*

everyone to ask “can you hear me now?” and the California Milk Advisory Board hopes we still all know that milk “does a body good.” Certainly, if I take someone else’s paper and put my name on it, or block-edit entire paragraphs from someone else’s work into my own without citing it, I have plagiarised. On the other hand, to suggest that one single-sentence definition has been plagiarised because of its proximal similarity to another seems a bit of a stretch. The fact that search engines now make it possible to search over 8 billion websites with a single mouse click certainly makes it more likely that some string of a dozen words might be found to exist in more than one place. Once upon a time, a small cohort of trained professionals wrote virtually all published material. It was prepared carefully, edited thoroughly, and it was considered intellectual property or work product – material for which people were paid. Under such circumstances, plagiarism should have been rare, and was rightly viewed as a form of theft. Are these still the circumstances of today?

Consider that there are well over a billion people using computers (Computer Industry Almanac Inc., n. d.). Also consider the explosive growth of social media. As of this writing, the US population is approximately 312 million, the population of India is approximately 1.2 billion, and the number of users of Facebook is approximately 800 million (Facebook Statistics, n. d.). So, while about 4.5% of the population of the world is American, nearly 11.5% of the population is on Facebook. Every one of those 11.5% are now potential producers of content, yet almost none of them are trained or paid professionals, and almost none of them undergo any form of editing or oversight. Is it so difficult to consider that ubiquity might become a factor?

Given that there is 1) a relatively fixed number of words in actual active use, 2) an enormous number of people using those words, and 3) a media and now increasingly even an education system that intentionally use the repetition of short phrases to guide our learning and our behaviour, I think it may be time to reconsider plagiarism. After all, St. Thomas Aquinas told us that *repeticio est mater studiorum* (repetition is the mother of learning), so it seems hypocritical to criminalise that learning. While it is no doubt a fair criticism that using ubiquitous speech is uncreative, and perhaps even uninteresting, neither of those criticisms rises to the level of intellectual theft. We have created a media environment in which we increasingly saturate ourselves with maddeningly repetitious sound bytes, catchphrases, and advertising jingles, an education environment that encourages rote memorisation, and a social environment that limits our exchanges to 140 characters typed with our thumbs. Since we are all products of those environments, it seems timely to reconsider how restrictively we define plagiarism.

About the author

Dr. E. Duff Wrobbel is Associate Professor of Communication at Southern Illinois University Edwardsville, where he has been on the faculty since 1993. His research interests include both conversation analysis and assessment. He is very involved in university governance issues, with particular interest in the relationship between governance structures and the quality of shared governance. His interest in plagiarism began when his campus became embroiled in a plagiarism scandal while he was President of his faculty Senate.

References

Computer Industry Almanac Inc. (n. d.) Retrieved from <http://www.c-i-a.com/>
Edlund, J. R. (n. d.). What is “plagiarism” and why do people do it? Retrieved from http://www.calstatela.edu/centers/write_cn/plagiarism.htm

- Facebook Statistics. (n. d.) Retrieved from <https://www.facebook.com/press/info.php?statistics>
- Falkofske, J. (2006). Is it plagiarism? Brief analysis of probability. Retrieved from <https://sctc.ims.mnscu.edu/shared/FacultyTutorials/MathematicsOfPlagiarism.pdf>
- Stinchcomb, J. B. (n. d.). Recommended student guidelines: Ethics and academic integrity. Retrieved from http://www.fau.edu/cdsi/pdfethics_student_guidelines.pdf
- Vocabulary. (n. d.). Retrieved 15 August, 2011, from Wikipedia <http://en.wikipedia.org/wiki/Vocabulary>
- Weinhold, W. (2009, January 29). SIU accused of copying plagiarism policy. *Daily Egyptian*. Retrieved from <http://www.siude.com/siu-accused-of-copying-plagiarism-policy-1.1318397>
- Zechmeister, E. B., Chronis, A. M., Cull, W. L., D'Anna, C. A., & Healy, N. A. (1995). Growth of a *functionally important lexicon*. *Journal of Reading Behavior*, 27(2), 201 – 212