

Beyond Turnitin: A Pedagogical Framework for Identifying Plagiarism in Student Writing

Nadya Yakovchuk

University of Leicester, UK.

Abstract

Effective plagiarism detection is one of the pillars of the holistic approach to addressing plagiarism. Specialised electronic detection software that can assess the degree of textual similarity of a piece of writing against a database of sources has been steadily gaining in popularity in recent years. Although electronic tools such as Turnitin UK offer wide scope of opportunity with regard to systematic screening of student work, they have built-in restrictions which do not allow them to serve as a 'panacea' for plagiarism identification.

This paper draws on a study of plagiaristic practices of undergraduate students at the University of Warwick, UK, and presents a Plagiarism Identification Framework developed in the course of this study. The elements of the Framework are described and the results of its application to student writing are presented, with reference to the Turnitin UK output received for the same samples of student writing. The paper considers the strengths and limitations of the two procedures for identifying plagiarism and discusses pedagogical implications of the proposed Plagiarism Identification Framework for EAP and subject tutors.

Introduction

Effective plagiarism detection is one of the pillars of the holistic approach to addressing plagiarism, the approach which, in recent years, has been continuously emphasised in the literature (Carroll and Appleton, 2001; Carroll, 2002; Freewood, Macdonald and Ashworth, 2003; Park, 2003; Duggan, 2006a). Effective detection implies the ability to correctly identify the presence and the amount of plagiarism in a certain piece of text. Identifying plagiarism, however, is not an easy task (Pennycook, 1994; Buranen, 1999; Angéilil-Carter, 2000; Park, 2003; Pecorari, 2003).

Specialised electronic detection software that can assess the degree of textual similarity of a piece of writing against a database of sources has been steadily gaining in popularity in recent years. Turnitin UK is the most well-known electronic detection service of this kind in the UK, with over 80 % of universities using it nationally (JISCiPAS, 2008). Apart from its main purpose of identifying textual similarity, the pedagogical potential of using Turnitin to educate students about plagiarism and proper use of sources has been also attracting increasing attention (Barrett, 2007; Haigh and Meddings, 2007; Peacock and Sharp, 2007).

Although electronic tools such as Turnitin offer wide scope of opportunity with regard to systematic screening of student work, they have built-in restrictions which do not allow them to serve as a 'panacea' for plagiarism identification. Although such tools are commonly referred to as *plagiarism detection software*, in fact they do *not* detect plagiarism, but only the level of *similarity* of a certain text with the sources in their databases. For example, Turnitin UK has a limited ability to determine whether 'unoriginal' passages are correctly cited and the Originality Reports it produces have to be checked carefully against the sources offered to determine the nature of identified textual similarities (Turnitin UK Instructor User Guide, 2005).

This paper draws on a study of plagiaristic practices of undergraduate students at the University of Warwick, UK, and presents a Plagiarism Identification Framework developed in the course of this study. The elements of the Framework are described and the results of its application to student writing are presented, with reference to the Turnitin UK output received for the same samples of student writing. The paper considers the strengths and limitations of the two procedures for identifying plagiarism and discusses pedagogical implications of the proposed Plagiarism Identification Framework for EAP and subject tutors.

Plagiarism Identification Framework

The creation of this Framework was inspired by Cambell's (1990), Pecorari's (2003) and Shi's (2004) studies of student textual borrowing practices. They proposed different schemes for coding student writing in terms of its reliance on sources, but none of them specifically aimed to identify plagiarism in student writing.

The Plagiarism Identification Framework developed in the course of this study was founded on three determinants for the acceptability of a written passage with regard to source incorporation:

- 1) Presence or absence of a *Quotation Signal* (QS) (either quotation marks for shorter quotations or the offset mode for longer quotations);
- 2) Presence or absence of a *Reference to a Source* (RS) (any form (i.e. author and date, number in brackets, footnotes, etc.) according to the referencing system used);
- 3) *Degree of Text Transformation*.

The last element, the *Degree of Text Transformation* had three levels: Exact Copying (EC), Wording Close to Original (WCO) which constituted an unacceptable paraphrase of the original text, and Wording Distant from Original (WDO) which represented acceptable paraphrase.

Various combinations of these three features (Quotation Signal, Reference to a Source and a Degree of Text Transformation) may or may not constitute acceptable writing in terms of source incorporation. This is illustrated in Table 1 which covers all the possible combinations of these features ('A' stands for 'Acceptable Writing', 'I' stands for 'Inconsistency in Source Attribution', and 'P' stands for 'Plagiarism').

Table 1. Three degrees of acceptability of writing

	Exact Copying	Wording Close to Original	Wording Distant from Original
QS -Yes RS - Yes	A	I	I
QS-Yes RS - No	I	I	I
QS – No RS - Yes	P	P	A
QS – No RS - No	P	P	P

This table represents the Plagiarism Identification Framework which was applied to ten samples of academic writing of third-year Chinese students studying for a BA in English Language, Translation and Cultural Studies at the University of Warwick, UK. The samples analysed in this study constituted literature review sections of their assignments prepared for the 'Sociolinguistics' course that I was teaching. Choosing the students and the area I was familiar with was a conscious decision which was very important for the study: being familiar with the main literature in the area I had a much better chance of detecting plagiarism in the writing of my students rather than in the writing of the students from a different department or on a different topic.

The average length of writing samples used for analysis was 1236 words (ranging from 421 to 1806). Each sample was divided into analysis units (mainly sentences), with the average

number of 54.8 units per sample (ranging from 24 to 88). Each of these units (548 in total) then had to be coded according to the Plagiarism Identification Framework presented above.

Results of Framework Application

Out of 548 analysis units, 123 were thought to be students' own writing, 298 were identified as based on sources, and 127 were coded as 'Unidentifiable' as it was not possible to ascertain their reliance on sources.

The results of the coding of 298 units that were based on sources are presented in Table 2.

Table 2. Percentage figures for three degrees of acceptability of writing

	Exact Copying	Wording Close to Original	Wording Distant from Original
QS -Yes RS - Yes	A (20.5)	I (0)	I (0)
QS-Yes RS - No	I (0)	I (0)	I (0.3)
QS – No RS - Yes	P (4)	P (12)	A (30.9)
QS – No RS - No	P (1.3)	P (10)	P (21)

As can be seen from this table, quite a considerable amount of plagiarism was identified in the analysis samples: 48.3% of the 298 analysis units that were based on sources (or 26.5% of all 548 analysis units across the ten samples).

One of the most notable findings was that 43% of the analysis units coded as 'Plagiarism' constituted plagiarism of ideas (WDO without a quotation signal or a source reference). The fact that this type of plagiarism accounts for such a substantial part of all the 'Plagiarism' units is a source of concern, particularly as plagiarism detection software, including Turnitin UK, cannot identify unacknowledged appropriation of ideas if a summary or a paraphrase is successful at a language level. A recent survey commissioned by *The Times Higher Education Supplement* also pointed to this problem: it found that copying ideas without acknowledgement is much more common among students than copying language (Shepherd, 2006).

29% of the 'Plagiarism' units constituted unacceptable paraphrase (WCO) with a source reference and without a quotation signal, i.e. the degree of text alteration at a language level was not enough for it to be considered a successful paraphrase. Since source acknowledgement was present for such units, rather than a deliberate attempt to present one's work as their own, this perhaps reflects student confusion over what constitutes acceptable and unacceptable paraphrase, which might have been further exacerbated by language difficulties (the writing was produced by students whose first language was not English). It is clearly one of the areas that need to be addressed pedagogically.

67% of 'Plagiarism' units did not contain a reference to a source. In the case of Exact Copies this could have signalled an intentional motive to deceive. However, in the case of WDO units (appropriate paraphrase) which constituted almost two thirds of the 'Plagiarism' units that did not have a reference to a source, this might also signal insufficient awareness of correct referencing conventions on the student's part: anecdotal evidence suggests that students often fall into the fallacy of thinking that if they make substantial changes to the original text then they do not have to acknowledge the source. This is another area that might require pedagogic attention.

Comparison with Turnitin UK Output

For reference purposes, the samples of student writing analysed in this study were also subjected to Turnitin UK screening. Six out of ten similarity indexes for these samples turned out to be 0%, three out of the remaining four which received either 1% or 2% similarity index were annulled after individual examination of each of these cases, which left only one sample with the index of 2%. In other words, according to Turnitin UK, these samples were, on the whole, plagiarism-free.

These results vary substantially from the results of applying the proposed Plagiarism Identification Framework to the analysis samples. As the Turnitin UK tool and the Framework-based procedure constitute very different approaches to plagiarism identification, these extensive differences do not allow for a straightforward comparison of the two procedures. In this study, however, the use of the Plagiarism Identification Framework, unlike the use of Turnitin UK, did highlight and bring to the fore a number of clearly problematic areas in student writing, many of which require appropriate pedagogical intervention. It is interesting to consider, therefore, how to position the two procedures in relation to each other.

Discussion

Turnitin UK offers a highly standardised procedure for identifying matching text which 'considerably reduces the amount of time required to track down sources of suspected plagiarism' (JISC PDS). This tool would probably work better for longer pieces of writing as it performs its operations fast, searching through billions of sources in a short period of time. In a way, it could act as a 'quick fix' for those academics that are in charge of large groups of students and want to check their writing for plagiarism. It would still require careful examining of each Originality Report on the part of academics to eliminate any unrelated findings, but, nevertheless, it would take significantly less time and effort doing this, rather than examining student writing manually.

The ability of this tool to perform its functions, however, is restricted due to a number of important limitations. Firstly, the Turnitin UK essay bank database, although quite extensive and under constant expansion, is limited, and its search covers only part of the Internet. Secondly, Turnitin UK does not carry out comparisons against printed sources (particularly books) which,

even with the increasing reliance on electronic sources, still constitute an important resource field for students and academics. In terms of the sources used by the subjects of my study, 49 out of 70 were paper-based and the remaining 21 were available electronically. As more paper-based resources are transferred into an electronic form, and subsequently included in the Turnitin UK database, the more powerful this tool will be in fulfilling its main function of identifying matching text. Thirdly, as mentioned earlier, Turnitin UK Originality Reports always need to be carefully examined by academics before making judgements regarding the amount of 'plagiarism' identified by this tool. As Ingram (2006: 2) pointed out, 'Artificial Intelligence as we all know is only as smart as the code in its program [...]. But at the moment the JISC Tool isn't that smart, so results still require close scrutiny'. And, finally, the Turnitin UK tool, as mentioned previously, cannot identify plagiarism of ideas.

Perhaps, an important role that Turnitin UK can play in the battle against plagiarism could be that of a deterrent. As JISC suggest themselves, 'although Originality Reports can be very effective at helping to identify suspected individual cases of plagiarism, JISC PDS plagiarism prevention works even more powerfully when used as a deterrent. Students who know that their work could come under effective scrutiny are much more likely to produce original work' (JISC Plagiarism Detection Service Instructor User Guide, 2005: 16). In fact, some educationists have already recognised the role of Turnitin UK in deterring plagiarism (Duggan, 2006b; Edwards and Ran, 2006; Ingram, 2006; Badge, Scott and Cann, 2007).

In comparison to Turnitin UK, the procedure of applying the proposed Plagiarism Identification Framework is open to the criticism of subjectivity and is less standardised. It can also be very time-consuming as it requires scrupulous manual examination of both student writing and external sources, which would not be realistic in the context of long assignments and large student groups. In order to be able to make informed judgements about the presence or absence of plagiarism in student work using the Framework, the investigator has to be familiar with the topic and the literature on the subject. These limitations would make it rather challenging to implement this Plagiarism Identification Framework in a wide variety of contexts.

However, the Framework could be very helpful when close examination of written text and background sources is instrumental to the investigator's task, i.e. when it is used as a pedagogical tool for developing the skills of effective academic writing and appropriate source incorporation. For example, within the context of teaching EAP (English for Academic Purposes), it could be very effective to use the Framework for structured writing tasks, i.e. when students are required to produce a piece of writing on a certain topic incorporating a number of pre-determined (usually not very long) sources into their writing. In such cases, a tutor would know the original sources very well and close manual examination of student writing using the Framework would not be too time-consuming and labour-intensive. Since the proposed Framework covers transmission of both language and ideas, it would allow the tutor to identify not only cases of

improper source acknowledgement, but also of effective and ineffective summaries and paraphrases. Such problematic areas can then be subjected to remedial pedagogical action either individually or in a group / class format using the same Framework.

The Framework can also be used by *subject* tutors who are concerned about their students' academic writing and referencing skills. It is common practice at some departments and institutions to set initial trial assignments at the beginning of a course, which would normally be short assignments that have to draw on restricted sources, and the 'assessment' and feedback students receive for such assignments would be for their personal use. Using the proposed Plagiarism Identification Framework at this stage could help tutors identify any problematic areas in their students' writing and raise students' awareness of correct source incorporation.

Conclusion

Plagiarism identification, as mentioned earlier, is a contentious area. Carroll and Appleton (2001: 4) pointed out that 'because the problem has been exacerbated by the impact of communication and information technology, it could be tempting to believe that a solution is provided through the same means, i.e. through electronic detection tools'. Electronic tools alone, however, cannot provide a perfect detection solution (and this paper has asserted this). In particular, some types of plagiarism (e.g. plagiarism of ideas) cannot be detected by such software due to its built-in restrictions.

The two procedures of plagiarism identification (Turnitin UK and the Plagiarism Identification Framework developed for this study) are quite different in many respects. Neither procedure can be considered as an absolutely reliable indicator of the amount of plagiarism in students' work. Perhaps, ideally, they could be used as complementary methods, dependant on the purposes and character of plagiarism screening. Also, both of them can be used as pedagogical tools to raise student awareness of the process of incorporating others' voices in their writing and, subsequently, to stimulate students to review their own work, making sure that it fulfills the academic requirements of correct source incorporation. The pedagogic potential of both procedures definitely needs to be explored further, particularly in the contexts of plagiarism prevention and the promotion of an academic culture of trust and integrity.

References

Angélie-Carter, S. 2000. *Stolen Language? Plagiarism in Writing*. UK, Essex: Pearson Education Limited.

Badge, J. L., Cann, A.J. and J. Scott. 2007. To cheat or not to cheat? A trial of the JISC Plagiarism Detection Service with biological sciences students. *Assessment and Evaluation in Higher Education*, 32/4, pp. 433-439.

Barrett, R. 2007. Embedding plagiarism detection software into practice and policy. In Duggan, F. (Ed.). *Proceedings of the 2nd International Plagiarism Conference*. Newcastle: Northumbria University Press, pp. 15-22.

Buranen, L. 1999. But I Wasn't Cheating: Plagiarism and Cross-Cultural Mythology. In Buranen, L. and A. M. Roy (Eds.). *Perspectives on Plagiarism and Intellectual Property in a Postmodern World*. New York: State University of New York Press.

Campbell, C. 1990. Writing with others' words: Using background reading text in academic compositions. In Kroll B. (Ed.). *Second Language Writing*. Cambridge: Cambridge University Press, pp. 211-230.

Carroll, J. 2002. *A Handbook for Deterring Plagiarism in Higher Education*. Oxford: Oxford Centre for Staff and Learning Development.

Carroll, J. and Appleton, J. 2001. *Plagiarism. A Good Practice Guide*. Oxford Brookes University.

Duggan, F. 2006 (a). Plagiarism: prevention, practice and policy. *Assessment & Evaluation in Higher Education*, 31/2, pp. 151-154.

Duggan, F. 2006 (b). Why do students plagiarise? *Update Newsletter*. Northumbria Learning. May 2006, p.3.

Edwards, V. and A. Ran. 2006. *Meeting the Needs of Chinese Students in British Higher Education*. Sino-British Fellowship trust report.

http://www.ncll.org.uk/10_about/50_research/10_research_projects/chinesestudents.html

(Access date: 18/02/2008).

Freewood, M., Macdonald, R. and P. Ashworth. 2003. Why merely *policing* plagiarism is not the answer. In Rust, D. (Ed.). *Improving Student Learning: Ten Years On. Proceedings of the 10th Improving Student Learning Symposium*. Oxford: Oxford Centre for Staff and Learning Development.

Haigh, J. and F. Meddings. 2007. Using the Turnitin plagiarism detection tool to promote academic integrity. *JISC PAS Case Study*.

Available at: http://www.jiscpas.ac.uk/documents/haighmeddings_casestudy.pdf (Access date: 06/02/2008).

Ingram, T. 2006. Delivering the JISC Plagiarism tool as a preventive measure to poor academic writing rather than a detection method (DETER). *Interactions Journal*, 10/27.

<http://www2.warwick.ac.uk/services/cap/resources/pubs/interactions/archive/issue27/ingram>

(Access date: 18/02/2008).

JISCiPAS. 2008. *Turnitin UK - Electronic Plagiarism Detection*.

<http://www.jiscpas.ac.uk/TurnitinUK.php> (Access date: 26/02/2008).

JISC Plagiarism Detection Service Instructor User Guide. 2005 (Updated).

http://www.submit.ac.uk/usage_jisc/jisc_instructor_guide.pdf (Access date: 28/02/2008).

Park, C. 2003. In Other (People's) Words: plagiarism by university students – literature and lessons. *Assessment & Evaluation in Higher Education*, 28/5, pp. 471-488.

Peacock, S. and J. Sharp. 2007. A holistic approach to deploying TurnitinUK that encourages sustainable change in institutional plagiarism practice. In Duggan, F. (Ed.). *Proceedings of the 2nd International Plagiarism Conference*. Newcastle: Northumbria University Press, pp. 237-238.

Pecorari, D. 2003. Good and original: Plagiarism and patchwriting in academic second-language writing. *Journal of Second Language Writing*, 12, pp. 317-345.

Pennycook A. 1994. The Complex Contexts of Plagiarism: A Reply to Deckert. *Journal of Second Language Writing*, 3/3, pp. 277-284.

Shepherd, J. 2006. 1 in 6 admits to cheating. *The Times Higher Education Supplement*, March 17.

Shi, L. 2004. Textual Borrowing in Second-Language Writing. *Written Communication*, 21/2, pp. 171-200.