
An exploration of the identification of the use paraphrasing tools and online language translation tools in student academic texts

Felicity M. PRENTICE, Clare E. KINDEN
La Trobe College Australia, Australia

The development and widespread implementation of word matching software programs to detect plagiarism has assisted in identifying episodes of academic misconduct (Atkinson & Yeoh, 2008; Sattler, Wiegel, & Veen, 2017). However as the battle metaphors in academic integrity suggest (Asefa & Coalter, 2007; Atkinson, Zaung Nau, & Symons, 2016; Crisp, 2007; Singh & Bennington, 2012), detecting new ways of breaching academic integrity may be viewed as an arms race. Students, under the pressures of time management, competing demands and the cultural dissonance experienced when their transactional approach to learning collides with the concepts of academic integrity espoused by Universities, will seek new ways to circumvent academic integrity (Bretag et al., 2018; Harper et al., 2018).

In this paper we describe our experiences when encountering student academic misconduct in an undergraduate Health Sciences unit of study. The assessment task required students to submit an essay in response to a specific case study prompt. A number of the essays exhibited significant similarity of content, suggesting collusion, however each essay was unique in the language used. The language stood out as not only unidiomatic, but also extraordinary in the use of bizarre synonyms for expected medical terminology.

When working with students who have English as an Additional Language (EAL), we have encountered some written work which is of a low standard of English expression. In these cases we often assumed that the essay had been initially written in the students' first language, and then put through an online translation tool such as Google Translate™ (<http://google.translate.com.au>), resulting in syntactical and semantic errors. Although this represents poor academic practice, it could be argued that as the original source had been the product of the student's intellectual endeavour, it is not strictly plagiarism (Mundt & Groves, 2016). Students, as entrants to the discourse of writing in the Health Sciences, have also demonstrated patchwriting, where segments of text may be appropriated from other sources and roughly rewritten to evade word matching software. Howard defines patchwriting as,

Copying from a source text and then deleting some words, altering grammatical structures, or plugging one synonym for another.

Howard, 1999, in Jamieson, 2015, p. xvii

While some synonym substitution may be employed in patchwriting, students in the Health Sciences are required and expected to use specific and standardised medical terminology. In the essays in question, synonyms were used extensively, including the substitution of lay terms for medical nomenclature.

Our curiosity was piqued when, in one essay, a student had directly transcribed and plagiarised from Wikipedia a description of a Computerised Axial Tomography scan (CAT Scan). However, in describing images taken from various angles, they had misspelled the word



'angles' as 'angels'. In a different student's essay the description of the CAT Scan declared that the images were taken from various 'Blessed Messengers'.

Work by Rogerson and McCarthy, 2017, had raised our awareness of students' use of online free paraphrasing tools to 'spin' text in such a way as to deceive word matching software such as Turnitin®, and in this case we believed that this technique may have been employed.

Paraphrasing tools were initially developed to 'spin', that is to generate, new and different versions of source text material to create and populate multiple websites. These websites would contain links to an original webpage, and thus improve the Google Search Index for that page. In this process, referred to as Black Hat Marketing, it is imperative that the multiple websites have text which cannot be identified by word matching software, as this would result in penalties imposed by Google (Lancaster, 2009). Students have subsequently employed these tools to spin text from websites, articles and other students' work to create 'plagiarism free' essays. The paraphrasing tools work by synonym substitution, however they lack the discretionary powers to analyse the re-engineered text for readability and comprehension by human readers. Our examination of the essays which we suspected had been subject to paraphrasing tools demonstrated a profound lack of readability, and the inclusion of synonyms for standardised medical terminology.

If the students had used paraphrasing tools on the work of other authors, it was clearly plagiarism. However, we were concerned that the students may have been subjecting their own work, prepared in a LOTE, to online language translation tools, and thus not breaching academic integrity.

We set out to investigate the question "Are there identifiable features which can differentiate text that has been processed by English-to-English paraphrasing tools from text processed by LOTE-to-English translation tools?" In particular we were looking for the presence of unidiomatic language and the use of synonyms for standardised medical terminology.

We selected a corpus of text from the essay case study prompt which featured standardised health sciences terminology and subjected it to iterative translation through six languages (Arabic, Punjabi, Hindi, Traditional and Simplified Chinese and Vietnamese) using Google Translate™, and through a six free online paraphrasing tools.

The results from Google Translate™ produced readable text, with minor errors in tense, verbal case matching and pronoun gender. Significantly, medical terminology was largely preserved by the language translation tool, with more synonym substitutions found in Vietnamese and Arabic, and less in Punjabi, Hindi and Chinese (Traditional and Simplified), while grammatical errors were more common in Arabic, Hindi and Punjabi.

The results demonstrated that free online paraphrasing tools did not identify medical terminology as standardised nomenclature and substituted a diverse range of often inappropriate synonyms. From the 21 standard medical terms in the source text the paraphrasing tools generated 73 synonyms, compared to only 7 synonyms from Google Translate™. The level of intelligibility of the text varied between paraphrasing tools, for example, from the original (seed) text:

One day, while Doug was out walking, he felt lightheaded and then lost consciousness and fell to the ground. He was brought to the Emergency Department of a major hospital by ambulance for assessment and investigation.

The following results were obtained:

Brace girl, stretch Doug was at large peripatetic, he felt lightheaded and fit lost consciousness and fell to the ground. He was debasement to the Danger Diversify of a chief sanatorium by ambulance for weight and criticism.

<http://plagiarisma.net/spinner.php>

One sidereal day, while Doug was out walk, he felt lightheaded and then lost know- ingness and downslope to the pulverization. He was brought to the Emergency Department of a major hospital by ambulance for assessment and probe.

<https://www.rephraser.net/instant-paraphrasing-tool/>

Further to this, one of the paraphrasing tool websites provided the suggestion that as the paraphrased text was so poor the student would do better to outsource the assignment to a third-party contract cheating site and provided a direct link to this service.

From this brief investigation we have increased confidence that it is possible to differenti- ate between online language translation and online paraphrasing tools, with the following criteria suggested:

Free online paraphrasing tools:

- the output is frequently of such poor quality as to render the text unintelligible,
- the language generated will be notable for the use of unidiomatic words and phrases,
- expected vocabulary such as standard medical terminology will usually be substituted with inappropriate synonyms,
- word matching software, such as Turnitin[®], may not recognise the re-engineered text from the source and thus a low similarity index may not be indicative of the level of plagiarism.

Online translation tools, such as Google Translate[™]:

- While there may be some anomalies in the language, overall the text will be intelligible,
- there is less likelihood that discipline specific nomenclature, such as standard medical terminology, will be changed to the same extent as paraphrasing tools.

Keywords: paraphrasing tools, online language translation tools, medical terminology, plagiarism.



References

Asefa, S., & Coalter, T. (2007). Factors that Influence Faculty Decisions Regarding Academic Dishonesty : A Study of a Private Liberal Arts University. *The Journal of Academic Administration in Higher Education*, 3, 43–51.

Atkinson, D., & Yeoh, S. (2008). Student and staff perceptions of the effectiveness of plagiarism detection software. *Australasian Journal of Educational Technology*, 24(2), 222–240.

Atkinson, D., Zaung Nau, S., & Symons, C. (2016). Ten Years in the Academic Integrity Trenches: Experiences and Issues. *Journal of Information Systems Education*, 27(3), 197–207. Retrieved from <http://esc-web.lib.cbs.dk/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=124558788&site=ehost-live&scope=site>

Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., Rozenberg, P., ... van Haeringen, K. (2018). Contract cheating: a survey of Australian university students. *Studies in Higher Education*, 0(0), 1–20. <https://doi.org/10.1080/03075079.2018.1462788>

Crisp, G. T. (2007). Staff attitudes to dealing with plagiarism issues: Perspectives from one Australian university. *International Journal for Educational Integrity*, 3(1), 3–15. Retrieved from <http://www.ojs.unisa.edu.au/journals/index.php/IJEI/>

Harper, R., Bretag, T., Ellis, C., Newton, P., Rozenberg, P., Saddiqui, S., & van Haeringen, K. (2018). Contract cheating: a survey of Australian university staff. *Studies in Higher Education*, 0(0), 1–17. <https://doi.org/10.1080/03075079.2018.1462789>

Lancaster, T. (2009). Automated Essay Spinning - An Initial Investigation 10 th Annual Conference of the Subject Centre for Information and Computer Sciences 25th – 27th August 2009 University of Kent at Canterbury Edited by : Hazel White , Higher Education Academy ICS Subject.

Mundt, K., & Groves, M. (2016). A double-edged sword: the merits and the policy implications of Google Translate in higher education. *European Journal of Higher Education*, 6(4), 387–401. <https://doi.org/10.1080/21568235.2016.1172248>

Sattler, S., Wiegel, C., & Veen, F. van. (2017). The use frequency of 10 different methods for preventing and detecting academic dishonesty and the factors influencing their use. *Studies in Higher Education*, 42(6), 1126–1144. <https://doi.org/10.1080/03075079.2015.1085007>

Singh, H., & Bennington, A. J. (2012). Faculty on the frontline: Predicting faculty intentions to address college student plagiarism. *Academy of Educational Leadership Journal*, 16(4), 115–128.

Note

This material presented in this extended abstract is currently under consideration for publication in the *International Journal of Educational Integrity*.