

'I didn't mean to steal someone else's words!': A Forensic Linguistic Approach to Detecting Intentional Plagiarism

Rui Sousa Silva^{1,2}, Tim Grant¹, Belinda Maia²

¹Aston University, Birmingham, United Kingdom, ²CLUP - Universidade do Porto, Porto, Portugal

The concept of plagiarism is not uncommonly associated with the concept of intellectual property, both for historical and legal reasons: the approach to the ownership of 'moral', non-material goods has evolved to the right to individual property, and consequently a need was raised to establish a legal framework to cope with the infringement of those rights. The solution to plagiarism therefore falls most often under two categories: ethical and legal. On the ethical side, education and intercultural studies have addressed plagiarism critically, not only as a means to improve academic ethics policies (PlagiarismAdvice.org, 2008), but mainly to demonstrate that if anything the concept of plagiarism is far from being universal (Howard & Robillard, 2008). Even if differently, Howard (1995) and Scollon (1994, 1995) argued, and Angèlil-Carter (2000) and Pecorari (2008) later emphasised that the concept of plagiarism cannot be studied on the grounds that one definition is clearly understandable by everyone. Scollon (1994, 1995), for example, claimed that authorship attribution is particularly a problem in non-native writing in English, and so did Pecorari (2008) in her comprehensive analysis of academic plagiarism. If among higher education students plagiarism is often a problem of literacy, with prior, conflicting social discourses that may interfere with academic discourse, as Angèlil-Carter (2000) demonstrates, we then have to aver that a distinction should be made between intentional and inadvertent plagiarism: plagiarism should be prosecuted when intentional, but if it is part of the learning process and results from the plagiarist's unfamiliarity with the text or topic it should be considered 'positive plagiarism' (Howard, 1995: 796) and hence not an offense. Determining the intention behind the instances of plagiarism therefore determines the nature of the disciplinary action adopted.

Unfortunately, in order to demonstrate the intention to deceive and charge students with accusations of plagiarism, teachers necessarily have to position themselves as 'plagiarism police', although it has been argued otherwise (Robillard, 2008). Practice demonstrates that in their daily activities teachers will find themselves being required a command of investigative skills and tools that they most often lack.

We thus claim that the 'intention to deceive' cannot inevitably be dissociated from plagiarism as a legal issue, even if Garner (2009) asserts that generally plagiarism is immoral but not illegal, and Goldstein (2003) makes the same severance. However, these claims, and the claim that only cases of copyright infringement tend to go to court, have recently been challenged, mainly by forensic linguists, who have been actively involved in cases of plagiarism. Turell (2008), for instance, demonstrated that plagiarism is often connoted with an illegal appropriation of ideas. Previously, she (Turell, 2004) had demonstrated by comparison of four translations of Shakespeare's *Julius Caesar* to Spanish that the use of linguistic evidence is able to demonstrate instances of plagiarism. This challenge is also reinforced by practice in international organisations, such as the IEEE, to whom plagiarism potentially has 'severe ethical and legal consequences' (IEEE, 2006: 57). What plagiarism definitions used by publishers and organisations have in common – and which the academia usually lacks – is their focus on the legal nature. We speculate that this is due to the relation they intentionally establish with copyright laws, whereas in education the focus tends to shift from the legal to the ethical aspects.

However, the number of plagiarism cases taken to court is very small, and jurisprudence is still being developed on the topic. In countries within the Civil Law tradition, Turell (2008) claims, (forensic) linguists are seldom called upon as expert witnesses in cases of plagiarism, either because plagiarists are rarely taken to court or because there is little tradition of accepting linguistic evidence. In spite of the investigative and evidential potential of forensic linguistics to demonstrate the plagiarist's intention or otherwise, this potential is restricted by the ability to identify a text as being suspect of plagiarism. In an era with such a massive textual production, 'policing' plagiarism thus becomes an extraordinarily difficult task without the assistance of plagiarism detection systems. Although plagiarism detection has attracted the attention of computer engineers and software developers for years, a lot of research is still

needed. Given the investigative nature of academic plagiarism, plagiarism detection has of necessity to consider not only concepts of education and computational linguistics, but also forensic linguistics. Especially, if intended to counter claims of being a 'simplistic response' (Robillard & Howard, 2008).

In this paper, we use a corpus of essays written by university students who were accused of plagiarism, to demonstrate that a forensic linguistic analysis of improper paraphrasing in suspect texts has the potential to identify and provide evidence of intention. A linguistic analysis of the corpus texts shows that the plagiarist acts on the paradigmatic axis to replace relevant lexical items with a related word from the same semantic field, i.e. a synonym, a subordinate, a superordinate, etc. In other words, relevant lexical items were replaced with related, but not identical, ones. Additionally, the analysis demonstrates that the word order is often changed intentionally to disguise the borrowing. On the other hand, the linguistic analysis of linking and explanatory verbs (i.e. referencing verbs) and prepositions shows that these have the potential to discriminate instances of 'patchwriting' and instances of plagiarism. This research demonstrates that the referencing verbs are borrowed from the original in an attempt to construct the new text cohesively when the plagiarism is inadvertent, and that the plagiarist has made an effort to prevent the reader from identifying the text as plagiarism, when it is intentional. In some of these cases, the referencing elements prove being able to identify direct quotations and thus 'betray' and denounce plagiarism. Finally, we demonstrate that a forensic linguistic analysis of these verbs is critical to allow detection software to identify them as proper paraphrasing and not – mistakenly and simplistically – as plagiarism.

(961 Words)

References

- Angèlil-Carter, S. (2000). *Stolen language? : plagiarism in writing*. Real Language Series. Harlow: Longman.
- Garner, B. A. (2009). *Black's Law Dictionary*. St. Paul, MN: West, 9th edition.
- Goldstein, P. (2003). *Copyright's highway: from Gutenberg to the celestial jukebox*. Stanford: Stanford University Press.
- Howard, R. (1995). *Plagiarisms, authorships, and the academic death penalty*. *College English*, 57(7), 788–806.
- Howard, R. & Robillard, A. Eds. (2008). *Pluralizing Plagiarism: Identities, Contexts, Pedagogies*, p. 1–6. Boynton/Cook: Portsmouth, 1 edition.
- IEEE, (2006). *IEEE Publication Services and Products Board Operations Manual*. New Jersey: IEEE Publications.
- Pecorari, D. (2008). *Academic Writing and Plagiarism: A Linguistic Anaylsis*. London: Continuum.
- PlagiarismAdvice.org, (2008). *3rd international plagiarism conference: Towards an authentic future*. Northumbria University.
- Robillard, A. E. (2008). *Situating plagiarism as a form of authorship: The politics of writing in a first-year writing course*. In R. Howard & A. Robillard, Eds., *Pluralizing Plagiarism: Identities, Contexts, Pedagogies*, p. 27–42. Portsmouth: Boynton/Cook, 1 edition.
- Robillard, A. E. & Howard, R. M. (2008). *Plagiarisms*. In R. M. Howard & A. E. Robillard, Eds., *Pluralizing Plagiarism: Identities, Contexts, Pedagogies*, p. 1–7. Boynton/Cook.
- Scollon, R. (1994). *As a matter of fact: The changing ideology of authorship and responsibility in discourse*. *World Englishes*, 13(1), 33–46.

Scollon, R. (1995). *Plagiarism and ideology: Identity in intercultural discourse*. *Language in Society*, 24, 1–28.

Turell, M. T. (2004). *Textual kidnapping revisited: the case of plagiarism in literary translation*. *The International Journal of Speech, Language and the Law*, 11(1), 1–26.

Turell, M. T. (2008). *Plagiarism*. In *Dimensions of Forensic Linguistics*, p. 265–99. Amsterdam / Philadelphia: John Benjamins Publishing Company.