

# Detecting Plagiarism in the Forensic Linguistics Turn



**Dr Rui Sousa-Silva's** research interests lie in **within-language** and **between-language** plagiarism, as well as in authorship analysis. He is a member of CLUP, the Linguistics Centre of the University of Porto, and of the Forensic Linguistics Group of UFSC, the Federal University of Santa Catarina (Brazil).

## The Research Questions

- (a) How do university students re-use text to plagiarise? (b) How does existing plagiarism detection software perform? Can software be designed to improve detection of different instances of plagiarism? (c) Can linguistic clues demonstrate the plagiarist's intention?
- (d) How can forensic linguistic analysis contribute to investigative and evidential detection?

## The Data

(1) Survey Data: Responses of students and lecturers/tutors from UK and Portugal to assess cross-cultural perceptions of plagiarism.

Pilot Survey				Final Survey			
Aston		Porto		Aston		Porto	
Lect.	Stud.	Lect.	Stud.	Lect.	Stud.	Lect.	Stud.
5	11	15	14	20	192	52	364

Survey Participants.

(2) Textual Data: Linguistic analysis of academic and non-academic texts of **CorRUPT** corpus (own Corpus of Reused and Plagiarised Texts) to identify strategies used to plagiarise.

## Analysis

(1) Statistically significant differences found in perceptions of students and lecturers/tutors of the two countries (UK and Portugal):

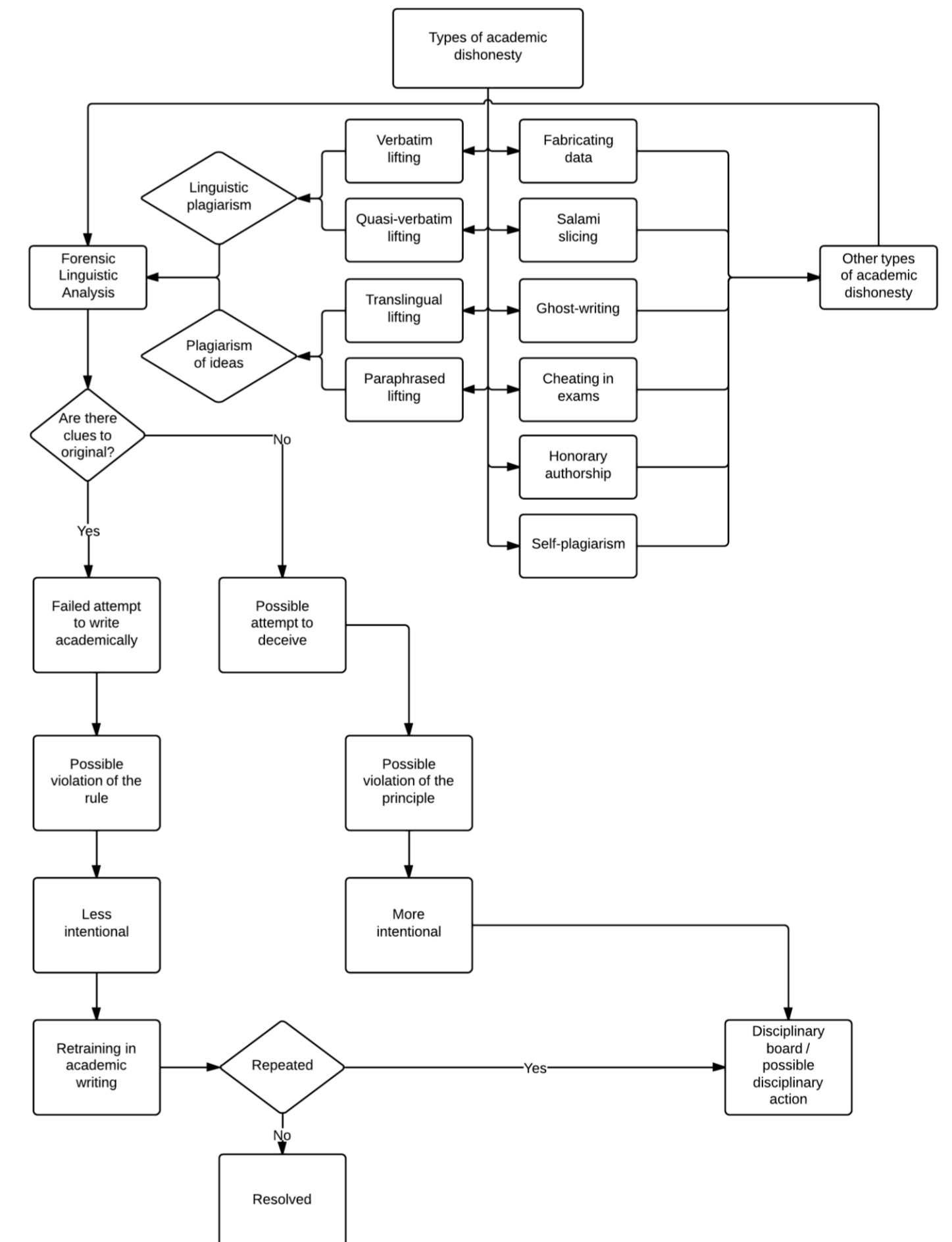
- (a) UK participants agree more with charges of plagiarism than Portuguese colleagues;
- (b) Students agree less strongly than lecturers/tutors with accusations of plagiarism;
- (c) Portuguese participants consider the plagiarist's intention more than UK colleagues.

(2) Main linguistic strategies used by plagiarists:

- (a) Lexical overlap;
- (b) Text reordering;
- (c) Word substitution and paraphrasing;
- (d) Translated text reuse.

## Findings and Implications

- (a) Further research required to improve automatic plagiarism detection.
- (b) Translated text reuse (**translingual plagiarism**) detectable linguistically.
- (c) Text provides clues to plagiarist's intention.
- (d) Plagiarism is not universally understood and varies cross-culturally.
- (e) Plagiarism is not only a moral, but also a legal problem.
- (f) Multidisciplinary forensic linguistic approach contributes to investigative and evidential plagiarism detection.
- (g) Computational forensic linguistics can reduce amount of manual work.



Plagiarism Handling Flowchart.