

# ‘I didn’t mean to steal someone else’s words!’: A Forensic Linguistic Approach to Detecting Intentional Plagiarism.

Rui Sousa-Silva,  
Tim Grant & Belinda Maia  
[sousasilva.rui@gmail.com](mailto:sousasilva.rui@gmail.com)

4th International  
Plagiarism Conference



# Plagiarism and Forensic Linguistics

## ▶ Plagiarism and Intellectual Property

- ▶ ethical aspects (e.g. Howard & Robillard, 2008; Angèlil-Carter, 2000; Pecorari, 2008)
- ▶ immoral vs illegal nature (e.g. Garner, 2009; Finnis, 1991; Eiras & Fortes, 2010)

## ▶ Intentional and inadvertent plagiarism

- ▶ ‘Positive plagiarism’ (Howard, 1995)

## ▶ Intentional vs. unintentional actions: criminal law

- ▶ Intention - doing something for which one does not have a lawful justification or excuse, that one knows will have those results
- ▶ Oblique intention – actions deriving from indirect intent, i.e. actions are not unintentional, but are not intended either; aka ‘culpable recklessness’: presupposes that the results derive, as a necessary consequence, from one’s actions
- ▶ Lack of intention (aka ‘un-intentionality’) - actions that happen by accident, or through mistakes or lack of foresight

# Intentional vs. Unintentional Plagiarism

## Intentional

- ▶ 'Virtually certain'
  - ▶ necessarily intended.
- ▶ The plagiarist knowingly uses the means to reach an end
  - ▶ knows accepted norms, i.e. the norms they are expected to conform to.
  - ▶ is aware of them, knows the causes implied.
  - ▶ chooses to plagiarise, or foresees and wants the result.
  - ▶ includes 'recklessness' or 'negligence'.

## Unintentional

- ▶ 'Virtually certain'
  - ▶ not necessarily intended.

– Criminal Law.

# What can Forensic Linguistics do?

## ▶ Comparative analysis

- ▶ determine valid, reliable textual evidence to demonstrate intention.
- ▶ consider different degrees of intention in plagiarism.
- ▶ assess textual manipulation.
- ▶ use textual analysis to identify linguistic patterns used by the writer to plagiarise.
- ▶ detect intention without further need for contextual information.
- ▶ e.g. paraphrasing and change of referencing verbs, while omitting the sources: intention to disguise authorship.

# Feature Selection

## ▶ Paraphrasing

- ▶ rewording (i.e., reusing the phrases of an author by expressing them in other words).
- ▶ text summarisation to express in the writer's words the ideas of the original.

## ▶ Referencing verbs

- ▶ verbs that are used to report (i.e., to reference) other people's findings.
- ▶ imply adopting a critical stance, knowing the direction of the text, and directing the reader of the text in a particular direction. (Pecorari, 2008)
- ▶ 'referencing verbs' preferred to 'reporting verbs': these are used to report results, one's own or other people's.

# Method of Analysis

## ▶ Finite *Corpus*

- ▶ PG student essays, average 3,000 words/essay.
- ▶ quantitative relevance: no correlation between the text size and the number of features to disguise authorship.
- ▶ analysis of instances showing changes – paraphrasing & referencing verbs.

Student	Words
S1	3,638
S2	1,370
S3	3,333
S4	4,629
S5	2,033

# Method of Analysis

## ▶ Comparative analysis

- ▶ written in Portuguese: morphologically more diverse and flexible.
- ▶ all essays considered by teachers to be plagiarised.
- ▶ teachers provided original assignments, indication of sections (verbatim) plagiarised and sources from which the text had been (putatively) borrowed.
- ▶ no plagiarism detection software used, as aim was to analyse textual strategies used by students to plagiarise.

# Method of Analysis

## ▶ Comparative analysis

- ▶ alignment of plagiarised assignments and original sources side-by-side.
- ▶ verbatim strings were discarded; analysis only of strings where effort was applied to introduce (minimal) changes.
- ▶ identification of the strings changed by paraphrase.
- ▶ identification of the instances where the referencing verbs of the original in-text citations were changed to make the text read as their own.
- ▶ analysis to determine whether the changes may have been introduced inadvertently or intentionally.

# Analysis and Results

- ▶ Replacing words on the paradigmatic axis with semantically-related words
  - ▶ replacing (lexical) items with words from the same semantic field (e.g. synonyms, antonyms, superordinates, co-hyponyms, etc).
  - ▶ difficult to detect by software – even more so with morphologically complex languages (e.g. Portuguese).
  - ▶ intentional in that writer **consciously** makes significant changes to the original.

## Original text

, opção que **condicionará** a forma e o conteúdo da publicação.

...os artistas do futurismo adotam como princípio **enxergar** a vida com olhos vorazes e febris...

...os futuristas tentaram **plasmear em** suas pinturas a **idéia** de dinamismo...

...alternância de planos, **superposição** de imagens fundidas ou encadeadas.

...o **fotodocumentalismo**...

## Suspect text

**Esta** opção **irá condicionar** a forma e o conteúdo da publicação.

...os artistas do Futurismo adoptam como princípio **olhar** a vida com olhos vorazes e febris...

...os futuristas tentaram **passar nas** suas pinturas a **ideia** de dinamismo...

...alternância de planos, a **sobreposição** de imagens fundidas ou encadeadas.

...o **fotodocumentário**...

## Original text

...um texto vulgar sobre **um cantor ou uma atriz**...

Será muito interessante confrontar o que pensam os alunos, **os professores e os pais** sobre as diversas questões da actualidade **educativa**.

...um repórter fotográfico pode estar a cobrir **um jogo de futebol** e, simultaneamente, a ir enviando as suas fotos digitais...

una frase em spray que decia "**Putá Burgos**"

## Suspect text

...um texto vulgar sobre **um designer ou um artista plástico**...

Será muito interessante confrontar o que pensam os alunos, **os profissionais e o público em geral** sobre as diversas questões da actualidade **cultural**.

...um repórter fotográfico pode estar a cobrir **a performance de um artista** e, simultaneamente, a ir enviando as suas fotos digitais...

uma frase em spray que diz "**Putá Espanha**"

# Analysis and Results

## ▶ Deleting specification

- ▶ i.e., deleting elements that give the text specification.
- ▶ making the text more general – and hence possibly more difficult to trace to its original sources.
- ▶ a case of oblique intention, rather than a case of intention: deleting elements from a text is often an editing process - an attempt to improve the text - rather than an intention to deceive.

## Original text

...jornalistas, desenhadores e fotógrafos, gráficos, responsáveis pela produção, vendedores e **um ou dois** directores...

Para além de incluir informações úteis para a comunidade escolar (**calendário escolar, actividades e reuniões agendadas, recomendações práticas, etc.**), o jornal...

É útil, num momento inicial, decidir — **num espectro que inclui o "fanzine", o jornal "oficial" e o suplemento de um jornal da terra** — qual o modelo...

## Suspect text

...jornalistas, desenhadores e fotógrafos, gráficos, responsáveis pela produção, vendedores e directores...

Para além de incluir informações úteis para a comunidade, o jornal...

É útil, num momento inicial, decidir qual o modelo de publicação que se considera mais interessante.

# Analysis and Results

- ▶ Adding specification
  - ▶ specification is added to the original to create the derivative text.
  - ▶ implies **knowingly** inserting text.
  - ▶ hence is more intentional than deleting specification.

---

## Original text

Se exceptuarmos a ideia de que o motivo deve surgir sempre centrado, grande parte destas regras mantem-se **na fotografia de notícias**.

## Suspect text

Se exceptuarmos a ideia de que o motivo deve surgir sempre centrado, grande parte destas regras mantêm-se **na fotografia de notícias, na fotografia noticiosa**.

# Analysis and Results

- ▶ Adding words to reused verbatim strings
  - ▶ detection systems can be misled by introducing in the verbatim-copied text new elements (e.g. spelling mistakes or text) that do not exist in the original.
  - ▶ differs from adding specification, which is an attempt to improve the text.
  - ▶ is more likely to be an intention to mislead.

## Original text

...a mesma Macworld de Fevereiro de 1997 já fazia o anúncio dos novas máquinas a 300Mhz...a relatividade do tempo acentua-se no fenómeno da sua compressão.

Com **vermelhos** berrantes, verdes intensos...

Mais: o fotojornalismo tem-se mesclado com a própria publicidade, como aconteceu nas campanhas da Benetton.

## Suspect text

a mesma Macworld de Fevereiro de 1997 já fazia o anúncio dos novas máquinas a 300Mhz...**no corrente ano de 2006 todos estes números já estão ligados ao passado...**a relatividade do tempo acentua-se no fenómeno da sua compressão.

Com **cores vermelhas** berrantes, verdes intensos...

Depois o fotojornalismo tem-se embaralhado com a própria publicidade, **e a fotografia publicitária**, como aconteceu nas campanhas da Benetton.

# Analysis and Results

- ▶ Omitting and replacing textual elements to make the derivative text coherent
  - ▶ the original is similar to the topic, but not exactly on it.
  - ▶ this feature demonstrates intention in that it requires a conscious stance from the writer to adapt the text to their aims.

## Original text

...órgão do clube de comunicação ou a publicação de um grupo de estudantes e professores.

...que indique, pelo menos, o título, a data de saída, o número, o director, os colaboradores (explicitando as suas tarefas), o estabelecimento de ensino, o endereço, o telefone, o e-mail, a periodicidade...

...do estabelecimento de ensino, as autarquias locais — Junta de Freguesia e Câmara Municipal — e o Instituto Português de Juventude são algumas das entidades...

## Suspect text

...órgão do clube de comunicação ou a publicação de um grupo de estudantes.

...que indique, pelo menos, o título, a data de saída, o número, o director, os colaboradores (explicitando as suas tarefas), a periodicidade...

(além naturalmente do as autarquias locais — Junta de Freguesia e Câmara Municipal — são algumas das entidades que podem apoiar o jornal).

# Analysis and Results

- ▶ Changing the spelling and morphology
  - ▶ Plagiarism is often disguised by changes in spelling, and is in that sense intentional.
  - ▶ Portuguese students are often said to plagiarise Brazilian texts, which requires changing the spelling in accordance with the EP conventions: e.g., where EP uses '-çç', '-pt' or '-ct', BP uses only '-ç', '-t' and '-t', respectively.
  - ▶ this straightforward and obvious change in spelling is sufficient to mislead automatic detection systems.
  - ▶ morphological changes also operate between variants, e.g. where BP tends to use a pronoun only, EP requires a determiner followed by a pronoun.

## Original text

...De grande repercussão social, **seus** princípios foram o ponto de partida para a modernização da cultura italiana.

Alguns jovens artistas tentaram reavivá-lo após 1918, mas sem sucesso; porém, **sua** influência sobre os outros movimentos modernos foi importante e duradoura.

**junto** com Giorgio De Chirico, **ele se separaria** finalmente do futurismo para se dedicar àquilo que eles próprios dariam o nome de Pintura Metafísica

## Suspect text

...De grande repercussão social, **os seus** princípios foram o ponto de partida para a modernização da cultura italiana.

Alguns jovens artistas tentaram reavivá-lo após 1918, mas sem sucesso, porém, **a sua** influência sobre os outros movimentos modernos foi importante e duradoura.

**juntamente** com Giorgio De Chirico, **separar-se-ia** finalmente do Futurismo para se dedicar àquilo que eles próprios dariam o nome de Pintura Metafísica

# Analysis and Results

- ▶ Changing the word order
  - ▶ it can be argued that simply changing the word order is not paraphrasing in the strict sense that the original words are not replaced with other words.
  - ▶ but it is paraphrasing in the sense that a new combination, containing the same words, is created.
  - ▶ passes undetected by most plagiarism detection systems.
  - ▶ prevents the source from being identified in search engines like Google, when searching for the exact strings.
  - ▶ changing the word order is intentional in that the meaning provided, as well as the individual words, are reused consciously without acknowledgement.

## Original text

...o Fauvismo e o Cubismo, **que foram assim chamados** por seus antagonistas, o Futurismo escolheu seu próprio nome

**No sentido lato, entendemos por fotojornalismo** a actividade de realização de fotografias informativas...

## Suspect text

...o Fauvismo e o Cubismo **assim foram chamados pelos** seus antagonistas, o Futurismo escolheu o seu próprio nome

**por fotojornalismo no sentido lato, entendemos** a actividade de realização de fotografias informativas...

# Analysis and Results

## Use of referencing verbs

- most referencing verbs in the *corpus* demonstrate a verbatim borrowing from the original, without further changes, e.g. the quoted text preceding or following the referencing verb.
- but the referencing verb is often reworked to create distance from the sources.
- whereas verbatim copying is more a case of recklessness, reworking the referencing verb requires a conscious attempt to change the original and is thus **intentional**.
- in these instances, the reused/adjusted referencing verbs were copied from other texts where the original references had been cited.

## Original text

"Tudo passa e o tempo foge", afirmara Begson.

...pois o que importa, **com** diz Marinetti, é...

...**sustentada por** Jesús González Requena...

## Suspect text

"Tudo passa e o tempo foge", afirmara Begson.

...pois o que importa, **com** diz Marinetti, é...

**Para** Jesus Requena...

# Features used in the corpus texts

Features/Writer	S1	S2	S3	S4	S5
1. Replacing words on the paradigmatic axis with semantically-related words	√	X	√	√	√
2. Adding specification	X	X	√	√	√
3. Deleting specification	√	X	√	√	√
4. Omitting and replacing textual elements to make the derivative text coherent	√	X	√	X	X
5. Adding words to reused verbatim strings	X	√	X	√	√
6. Changing the spelling and morphology	X	X	X	√	√
7. Changing the word order	X	X	X	√	√
8. Use of referencing verbs:	X	X	√	√	√
8a. Verbatim use	X	X	√	√	0
8b. Use with changes	X	X	0	0	√

## Conclusions

- ▶ Paraphrasing and referencing verbs perform differently in demonstrating the writer's intention.
- ▶ On balance, it can be thought that **using paraphrasing** and **borrowing referencing verbs** without acknowledgement can demonstrate **intention** or **oblique intention**.
- ▶ Intention includes:
  - ▶ replacing words on the paradigmatic axis with semantically-related words;
  - ▶ adding specification;
  - ▶ omitting and replacing textual elements to make the derivative text coherent;
  - ▶ adding words to reused verbatim strings, changing the spelling and morphology;
  - ▶ changing the word order;
  - ▶ introducing changes to referencing verbs.

## Conclusions

- ▶ Oblique intention includes:
  - ▶ deleting specification; and
  - ▶ verbatim using referencing verbs.
- ▶ Text manipulation equates with intention
  - ▶ not accidental or inadvertent.
  - ▶ requires awareness of conventions.
  - ▶ as referencing verbs imply adopting a particular stance and/or direct the reader in a particular direction: may indicate plagiarism or lack of academic writing skills.
  - ▶ borrowing mostly *verbatim* from the original – so should be treated like verbatim plagiarism.
  - ▶ intention to plagiarise is indicated by cases where the referencing verb is manipulated or replaced with a preposition (create distance from sources).

## Limitations and Future Work

- Present research that textual evidence can be used to demonstrate intention is still exploratory.
- Further tests, e.g. over a larger corpus necessary to argue for its generalisability.
- Study contributes to demonstrating intention in cases where teachers/supervisors may hesitate to pass judgment on the reuse of the original text by students without necessarily having to consider contextual information.

# References

- Angèlil-Carter, S. (2000). *Stolen language? : plagiarism in writing*. Real Language Series. Harlow: Longman.
- Eiras, H. & Fortes, G. (2010). *Dicionário de Direito Penal e Processo Penal*. Lisboa: Quid Juris.
- Finnis, J. (1991). 'Intention and side-effects'. In R. Frey & C. Morris, Eds., *Liability and responsibility: Essays in law and morals*, chapter 2, p. 32–64. Cambridge: Cambridge University Press.
- 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*. Boynton/Cook: Portsmouth.
- Pecorari, D. (2008). *Academic Writing and Plagiarism: A Linguistic Analysis*. London: Continuum.
- PlagiarismAdvice.org, (2008). '3rd international plagiarism conference: Towards an authentic future'. Northumbria University.
- Robillard, A. E. & Howard, R. M. (2008). 'Plagiarisms'. In R. M. Howard & A. E. Robillard, Eds., *Pluralizing Plagiarism: Identities, Contexts, Pedagogies*. 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. (2008). 'Plagiarism'. In *Dimensions of Forensic Linguistics*. Amsterdam / Philadelphia: John Benjamins Publishing Company.

This work was partially supported by grant SFRH/BD/47890/2008 FCT-Fundação para a Ciência e Tecnologia, Portugal, co-financed by POPH/FSE.



UNIÃO EUROPEIA  
Fundo Social Europeu



QUADRO  
DE REFERÊNCIA  
ESTRATÉGICO  
NACIONAL  
PORTUGAL 2007.2013



CLUP Centro de  
Linguística da  
Universidade do  
Porto

R&D Unit - POCI 2010