

Detecting translilingual plagiarism:

A forensic linguistic contribution to computational processing.

PAN @ CLEF 2016
Universidade de Évora
Évora, Portugal

Rui Sousa-Silva
Universidade do Porto.
r.sousa-silva@sapo.pt
6 september 2016



German minister Annette Schavan quits over 'plagiarism'

🕒 9 February 2013 | Europe

German defence minister resigns in PhD plagiarism row

Karl-Theodor zu Guttenberg had been stripped of doctorate by University of Bayreuth

World | Tue Dec 16, 2014 9:02am EST

Romanian PM gives up doctorate after years of plagiarism allegations

BUCHAREST | BY RADU-SORIN MARINAS

CORRECTING THE RECORD

CORRECTING THE RECORD; Times Reporter Who Resigned Leaves Long Trail of Deception

This article was reported and written by Dan Barry, David Barstow, Jonathan D. Glater, Adam Liptak and Jacques Steinberg. Research support was provided by Alain Delaquérière and Carolyn Wilder.
Published: May 11, 2003

Johann Hari leaves the Independent after plagiarism storm

Jonah Lehrer's Journalistic Misdeeds at Wired.com

An investigation reveals evidence of plagiarism, dodgy quotes, and factual inaccuracies.





(Adapted from http://www.insanelymac.com/forum/uploads/monthly_08_2012/post-364799-0-14742000-1345205066.jpg)



What is plagiarism?

- › Using WORDS without acknowledgement?
- › Using IDEAS without acknowledgement?
- › Using WORKS without acknowledgement?
- › Intentionally or unintentionally?

Plagiarism and Forensic Linguistics

- › The type of plagiarism that linguists are mostly competent to deal with is linguistic plagiarism (Coulthard & Johnson, 2007).
- › Plagiarism, language and the law.
- › Forensic linguistics expert: establish if two or more texts were:
 - a) produced independently,
 - b) based on the other (as a result of collusion), or
 - c) both derived from a prior text.
- › ‘Forensic’ as that which is of interest to the ‘forum’ (Turell, 2013).
- › Forensic Linguistics as an interdisciplinary field.
- › Plagiarism vs. text-matching.

Plagiarism is intentionally or knowingly reusing someone else's words, works or ideas without a proper, clear and unambiguous acknowledgement.

(Sousa-Silva, 2013)

Plagiarism and translation

- › Plagiarism in translated works (Turell, 2004): analysis of four Spanish translations of Shakespeare's *Julius Caesar* in Spanish.
- › Analysis of lexical overlap, shared once-only words, unique vocabulary and shared once-only phrases.
- › Concluded that no other text showed such an identical overlap as the suspect text.
- › A different case: **Translingual Plagiarism** (Sousa-Silva, 2013): text in the target language plagiarises text in another source language.



Translingual plagiarism

- › ‘Trans’: meaning *though, across*.
- › Interlanguage: an intermediate language that has an impact on the surface structure of a text.
- › Translingual plagiarism is comparatively under-researched:
 1. Research into plagiarism has been mostly English-centred.
 2. Offer / demand: most Internet is in English.
 3. Much more effort / data required to assist research into other languages.
- › Theoretical and methodological challenges to translingual plagiarism detection:
 1. How to detect plagiarism of texts translated from other languages?
 2. How to establish the comparison?
 3. How to explain and justify the findings?



Translingual plagiarism

- › In cases where machine translation is used to plagiarise, revert the procedure:
 1. The suspect text is translated into the language of the original available;
 2. a comparison of both texts is then made in order to establish whether the suspect text is an instance of plagiarism.

Em busca do autobronzeador ideal

Um creme que estimula as células produtoras do pigmento melanina em vez de simplesmente pintar a pele está a ser desenvolvido nos EUA.

O bronzeado do Verão dentro de uma garrafa é o que já prometem muitos produtos de beleza, mas os resultados nem sempre são os prometidos. Há os que deixam a pele cor-de-laranja, delineando uma clara fronteira de cor com a parte do corpo em que não foi esfregado, outros que fazem com que se tenha um ar de quem não se lava há um mês... Mas agora os cientistas inventaram o que promete ser o autobronzeador ideal e que, ainda por cima, protege contra o cancro da pele.

Por ora, as experiências foram feitas apenas com ratinhos sem pêlo, mas a equipa de John D'Orazio, da Universidade do Kentucky (EUA), relatou na revista "Nature" ter criado um creme que estimula as células a produzir o efeito de bronze, mas sem ser necessária a exposição ao Sol.

Até um certo limite de exposição, a luz solar pode deixar-nos com uma cor bonita. Mas, passados esses limites, a radiação pode provocar alterações no ADN ce-

lular que conduzam ao desenvolvimento de cancro. É contra isso que pretende lutar a equipa de D'Orazio: "Não estamos a fazer isto por motivos cosméticos – estamos a tentar encontrar formas de prevenir o cancro", explicou D'Orazio à revista "Wired".

"O melanoma é o cancro cuja incidência mais tem crescido nas últimas três décadas." O problema é não só a exposição à luz solar na praia, mas também a procura de um bronzeado rápido obtido sob as lâmpadas de um centro de estética. "Se este creme poder ajudar a evitar que as pessoas se deitem nas camas de bronzeamento, ficarei feliz", comentou.

A chave deste novo autobronzeador está num extracto de plantas chamado forskolina que, nas experiências da equipa, protegeu ratinhos sem pêlo de radiação ultravioleta e permitiu-lhes desenvolver um bronzeado natural, estimulando os seus melanócitos, que são células produtoras de pigmentos – que nos fazem ficar mais escuros, quando nos expomos ao Sol.

A capacidade de se bronzear – ou de simplesmente se queimar,

como acontece a algumas pessoas – é controlada pela hormona de estimulação dos melanócitos, que se liga a uma proteína que existe no exterior destas células. Esta proteína, que se chama receptor de melanocortina 1, funciona mal em muitas pessoas que têm a pele clara e o cabelo ruivo. É por isso que não conseguem bronzear, e ainda por cima correm maiores riscos de desenvolver cancro da pele.

A equipa de D'Orazio usou ratinhos que tinham esse receptor defeituoso, e concluiu que, se fossem tratados com o creme com forskolina, bronzeavam-se, sem ficarem queimados. Numa segunda série de experiências os cientistas usaram ratinhos susceptíveis ao cancro, expondo-os ao equivalente a uma a duas horas de Sol na altura do meio-dia solar, diariamente, durante 20 semanas. Os ratinhos que usaram o creme protetor desenvolveram menos tumores.

"Não há nenhuma razão para que este tratamento não resulte em seres humanos", comentou D'Orazio, citado pela revista "New Scientist". No entanto, este produto nunca foi testado em pessoas, por isso não se sabe se será seguro e tão eficiente como nos ratinhos, embora haja já algumas empresas interessadas em fazer ensaios com vista à sua comercialização. •

Os efeitos de UVA e UVB

Há frequências da radiação ultravioleta que bronzeiam a pele: UVA (315 a 400 nanometros) e UVB (280 a 315 nanometros).

Efeitos da UVA

- Faz com que os melanócitos libertem melanina que já está produzida.
- Faz com que a melanina se combine com oxigénio, o que produz o escurecimento da pele.
- Está relacionada com o rápido envelhecimento da pele.
- Parece ser menos cancerígena que a UVB, mas causa melanoma, que é uma forma perigosa de cancro da pele.
- Não é bloqueada pela maior parte dos protetores solares, mas pode ser travada, em boa parte, pela roupa

Efeitos da UVB

- Desencadeia a produção de mais melanina na pele.
- Distorce e danifica as cadeias de ADN das células.
- Causa o crescimento de sinais e algumas formas de cancro da pele (mas não melanoma).
- Envelhece a pele (menos que a UVA).
- Estimula a produção de vitamina D, que é essencial ao organismo, e diminuiu os riscos de outros cancos.
- Queima mais facilmente que a UVA se houver sobreexposição; em pequenas quantidades, é benéfica.
- Pode ser quase completamente bloqueada pelos protetores solares.





Did the journalist actually plagiarise?

The case against *Público*

PT	Faz com que a melanina se combine com oxigénio, o que produz o escurecimento da pele.
PT-EN	causes the melanin to combine with oxygen, which causes darkening of the skin
EN	causes the melanin to combine with oxygen (oxidize), which creates the actual tan color in the skin

Table 1

Example of translingual plagiarism.
(Adapted from my thesis.)



The case against *Público*

PT	Pode ser quase completamente bloqueada pelos protectores solares.
PT-EN	It can be almost completely blocked by sunscreen
EN	is almost completely blocked by virtually all sunscreens

Table 2

Example of translingual plagiarism.
(Adapted from my thesis.)



The case against *Público*

PT	A chave deste novo autobronzeador está num extracto de plantas chamado forskolina que, nas experiências da equipa, protegeu ratinhos sem pelo de radiação ultravioleta e permitiu-lhes desenvolver um bronzeado natural, estimulando os seus melanócitos (...)
PT-EN	<p>The key to this new self-tanning is a plant extract called forskolin that the experience of the team, protected hairless mice to ultraviolet radiation and allowed them to develop a natural tan by stimulating their melanocytes (...)</p>
EN	<p>The key chemical, a plant extract called forskolin, protected mice against UV rays and allowed them to develop a natural tan by stimulating pigment-producing cells called melanocytes.</p>

Table 3

Example of translational plagiarism.
(Adapted from my thesis.)

The case against *Público*

PT	A capacidade de se bronzear - (...) - é controlada pela hormona de estimulação dos melanócitos, que se liga a uma proteína que existe no exterior destas células. Esta proteína, que se chama receptor de melanocortina 1, funciona mal em muitas pessoas que têm a pele clara e o cabelo ruivo. É por isso que não se conseguem bronzear, e ainda por cima correm maiores riscos de desenvolver cancro de pele.
PT-EN	<p>The ability to tan - (...) - is controlled by hormone stimulation of melanocytes, which binds to a protein that exists outside these cells. This protein, called the melanocortin 1 receptor, malfunctions in many people who have fair skin and red hair. That is why we can not tan, and on top of a higher risk of developing skin cancer.</p>
EN	<p>The ability to tan is largely controlled by a hormone called melanocyte-stimulating hormone, which binds to the melanocortin 1 receptor (MC1R) on the outside of melanocytes. Many people with red hair and fair skin have a defect in this receptor, meaning they find it almost impossible to tan and are prone to skin cancer.</p>

Table 4

Example of
translingual plagiarism.
(Adapted from my thesis.)

The case against *Público*

PT	Numa segunda série de experiências os cientistas usaram ratinhos susceptíveis ao cancro, expondo-os ao equivalente a uma a duas horas de Sol na altura do meio-dia solar, diariamente, durante 20 semanas.
PT-EN	In a second series of experiments the scientists used mice susceptible to cancer, exposing them to the equivalent of one to two hours of sunshine at the time of solar noon each day for 20 weeks.
EN	In a second experiment, a particularly cancer-prone strain of mice, also bred to lack effective MC1Rs, were exposed to the equivalent of 1 to 2 hours of midday Florida sunlight each day for 20 weeks.

Table 5

Example of
translingual plagiarism.
(Adapted from my thesis.)

ador ideal

células produtoras
vez de simplesmente
envolvido nos EUA.

ular que conduzam ao desenvolvimento de cancro. É contra isso que pretende lutar a equipa de D’Orazio: “Não estamos a fazer isto por motivos cosméticos – estamos a tentar encontrar formas de prevenir o cancro”, explicou D’Orazio à revista “Wired”.

“O melanoma é o cancro cuja incidência mais tem crescido nas últimas três décadas.” O problema é não só a exposição à luz solar na praia, mas também a procura de um bronzeado rápido obtido sob as lâmpadas de um centro de estética. “Se este creme poder ajudar a evitar que as pessoas se deitem nas camas

como acontece a algumas pessoas – é controlada pela hormona de estimulação dos melanócitos, que se liga a uma proteína que existe no exterior destas células. Esta proteína, que se chama receptor de melanocortina 1, funciona mal em muitas pessoas que têm a pele clara e o cabelo ruivo. É por isso que não conseguem bronzear, e ainda por cima correm maiores riscos de desenvolver cancro da pele.

A equipa de D’Orazio usou ratinhos que tinham esse receptor defeituoso, e concluiu que, se fossem tratados com o creme com forskolina, bronzeavam-se, sem ficarem queimados. Numa segunda série de experiências os cientistas usaram ratinhos susceptíveis ao cancro, expondo-os ao equivalente a uma a duas horas de Sol na altura do meio-dia solar, diariamente, durante 20 semanas. Os ratinhos que usaram o creme protector desenvolveram menos tumores.

“Não há nenhuma razão para

Os efeitos de UVA e UVB

Há frequências da radiação ultravioleta que bronzeiam a pele: UVA (315 a 400 nanometros) e UVB (280 a 315 nanometros).

Efeitos da UVA

- Faz com que os melanócitos libertem melanina que já está produzida.
- Faz com que a melanina se combine com oxigénio, o que produz o escurecimento da pele.
- Está relacionada com o rápido envelhecimento da pele.
- Parece ser menos cancerígena que a UVB, mas causa melanoma, que é uma forma perigosa de cancro da pele.
- Não é bloqueada pela maior parte dos protectores solares, mas pode ser travada, em boa parte, pela roupa

is present more uniformly throughout the day, and throughout the seasons than UVB.

Efeitos da UVB

- Desencadeia a produção de

At its simplest, plagiarism, or more accurately the type of plagiarism we as linguists are competent to deal with, is the theft, or unacknowledged use, of text created by another.



At its simplest, plagiarism, or more accurately the type of plagiarism we as linguists are competent to deal with, is the theft, or unacknowledged use, of text created by another.

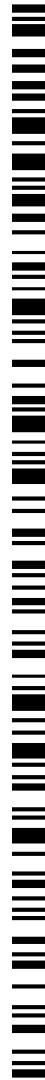
En su forma más simple, el plagio, o más exactamente el tipo de plagio nosotros, como lingüistas son competentes para tratar, es el robo o uso no reconocida, de texto creado por otro.



At its simplest, plagiarism, or more accurately the type of plagiarism we as linguists are competent to deal with, is the theft, or unacknowledged use, of text created by another.

En su forma más simple, el plagio, o más exactamente el tipo de plagio nosotros, como lingüistas son competentes para tratar, es el robo o uso no reconocida, de texto creado por otro.

Na sua forma mais simples, o plágio, ou mais exatamente o tipo de plágio nós, como linguistas são competentes para lidar com, é o roubo ou utilizar texto reconhecido criado por outro.



At its simplest, plagiarism, or more accurately the type of plagiarism we as linguists are competent to deal with, is the theft, or unacknowledged use, of text created by another.

Dans sa forme la plus simple, le plagiat, ou plus exactement le genre de plagiat nous en tant que linguistes sont compétents pour traiter, il est le vol ou utiliser le texte reconnu créé par d'autres.

En su forma más simple, el plagio, o más exactamente el tipo de plagio nosotros, como lingüistas son competentes para tratar, es el robo o uso no reconocida, de texto creado por otro.

Na sua forma mais simples, o plágio, ou mais exatamente o tipo de plágio nós, como linguistas são competentes para lidar com, é o roubo ou utilizar texto reconhecido criado por outro.

At its simplest, plagiarism, or more accurately the type of plagiarism we as linguists are competent to deal with, is the theft, or unacknowledged use, of text created by another.

Dans sa forme la plus simple, le plagiat, ou plus exactement le genre de plagiat nous en tant que linguistes sont compétents pour traiter, il est le vol ou utiliser le texte reconnu créé par d'autres.

En su forma más simple, el plagio, o más exactamente el tipo de plagio nosotros, como lingüistas son competentes para tratar, es el robo o uso no reconocida, de texto creado por otro.

Na sua forma mais simples, o plágio, ou mais exatamente o tipo de plágio nós, como linguistas são competentes para lidar com, é o roubo ou utilizar texto reconhecido criado por outro.

In its simplest form, plagiarism, or more precisely the sort of plagiarism us as linguists are competent to handle, it is the flight or use the recognized text created by others.

At its simplest, plagiarism, or more accurately the type of plagiarism we as linguists are competent to deal with, is the theft, or unacknowledged use, of text created by another.

In its simplest form, plagiarism, or more precisely the sort of plagiarism us as linguists are competent to handle, it is the flight or use the recognized text created by others.

A case of student plagiarism

- › Translingual plagiarism in a student's assignment.



CorRUPT

- › My own Corpus of Reused and Plagiarised Texts.



A case of student plagiarism

PT	Os procedimentos são baseados no corpo geralmente aceite de conhecimento e experiência no campo da examinação de documentação forense.
PT-EN	The procedures are based on generally accepted body of knowledge and experience in the field of forensic documentation examination.
EN	The procedures outlined here are grounded in the generally accepted body of knowledge and experience in the field of forensic document examination.

Table 6

Example of
translingual plagiarism.

A case of student plagiarism

PT	Um examinador forense pode confiavelmente chegar a um parecer sobre se dois ou mais itens manuscritas foram escritos pela mesma pessoa.
PT-EN	A forensic examiner can reliably reach an opinion on whether two or more handwritten items were written by the same person.
EN	By following these procedures, a forensic document examiner can reliably reach an opinion concerning whether two or more impressions were created by the same checkwriter, [...]

Table 7

Example of
translingual plagiarism.

A case of student plagiarism

PT	Neste método existem várias interferências, tais como:
PT-EN	In this method, there are various interferences such as:
EN	The following are limitations that may be present in a handwriting examination:

Table 8

Example of
translingual plagiarism.

A case of student plagiarism

PT	Os produtos apresentados para exame pode ter limitações inerentes que podem interferir com os procedimentos deste;
PT-EN	Products submitted for examination may have inherent limitations that may interfere with the procedures of this;
EN	Documents submitted for examination may have inherent limitations that can interfere with the procedures in this test method.

Table 9

Example of
translingual plagiarism.

A case of student plagiarism

PT	Existência de limitações devido a apresentação de documentos que não sejam originais, quantidade limitada ou comparabilidade, ou condição dos itens apresentados para exame;
PT-EN	Existence of limitations due to production of documents which are not original, limited quantity or comparability, or condition of the items submitted for examination;
EN	The generation of the document(s), limited quantity or comparability, or condition of the document(s) submitted may restrict examinations.

Table 10

Example of
translingual plagiarism.

A case of student plagiarism

PT	Mau armazenamento dos documentos, manipulação, teste ou processamento químico podem interferir com a capacidade de o examinador ver certas características. Sempre que possível, os exames de documentos devem ser realizados antes de qualquer tratamento químico. Os itens devem ser tratados de forma adequada para evitar comprometer exames subsequentes;
PT-EN	Bad storage of documents, handling , testing , or chemical processing can interfere with the ability of the examiner to view certain characteristics. Whenever possible, the document tests should be performed before any chemical treatment. Items must be handled appropriately to avoid compromising subsequent examinations ;
EN	The results of earlier storage , handling , testing , or chemical processing (e.g., latent prints and DNA) may interfere with the document examination. When possible, document examinations should be conducted before chemical processing. Documents should be handled properly to avoid compromising subsequent examinations .

Table 11

Example of
translingual plagiarism.

Another case of student plagiarism

- › A case of foreign students attempting to write in Portuguese.

Another case of student plagiarism

PT	Pois bem, se numa parede há uma frase em spray que diz "Puta Espanha", isso não é um graffiti, é uma pintada.
PT-ES	Bueno, si una pared es una frase aerosol que dice "Santo España", esto no es un graffiti está pintado .
ES	Pues bien, hace años, debajo de mi casa, había en una pared una frase en spray que decía "Puta Burgos" y eso no era un graffiti , era una pintada .

Table 12

Example of
translingual plagiarism.

Another case of student plagiarism

PT	Agora bem, em quase todas as cidades, há paredes com inscrições e desenhos por trás das quais há algo mais que o insulto a uma equipa de futebol, ou que um mural "bonito" pintado calmamente a plena luz do dia à cara dos transeúntes ou que o encarrego do rótulo de um negócio.
PT-ES	Ahora, en casi todas las ciudades, hay paredes con inscripciones y dibujos detrás de la cual hay algo más de sal a un equipo de fútbol, o una pared "hermosa" pintado en silencio plena luz del día en la cara de los transeúntes o la etiqueta del orden de un negocio.
ES	Ahora bien, cerca de mi casa otra vez (¡Vaya! Cerca de mi casa pasan cosas muy interesantes) hay una zona de paredes con inscripciones y dibujos detrás de las cuales hay algo más que el insulto a un equipo de fútbol, o que un mural "bonito" pintado tranquilamente a plena luz del día a la cara de los transeúntes o que el encargo del rótulo de un negocio.

Table 13

Example of
translingual plagiarism.

Another case of student plagiarism

PT	Há umas letras, às vezes uma personagem, às vezes um ícono...
PT-ES	Hay cartas, <i>a veces un</i> carácter, <i>a veces un</i> ícono ...
ES	Hay unas letras, <i>a veces un</i> personaje, <i>a veces un</i> ícono...

Table 14

Example of
translingual plagiarism.

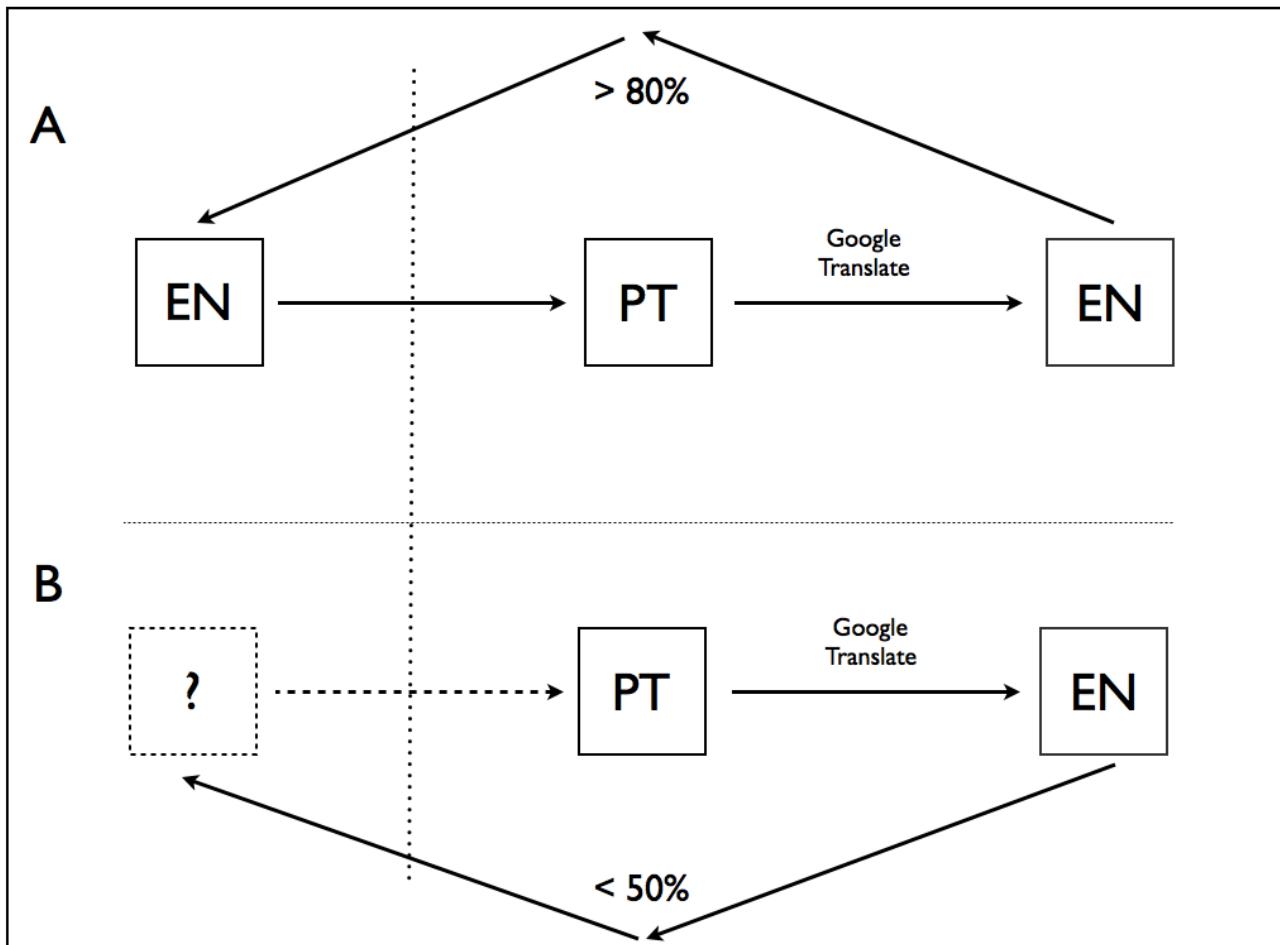
Another case of student plagiarism

PT	Não parecem nomes correntes parecem apodos, e nenhum deles coincide.
PT-ES	No parecen nombres comunes parecen apodos, y ninguno de ellos coinciden.
ES	No parecen nombre corrientes parecen apodos, y ninguno de ellos coincide.

Table 15

Example of
translingual plagiarism.

Investigating non-suspect texts



Figure

Diagram of translingual plagiarism detection.
(Adapted from my thesis.)

Investigating non-suspect texts

President Barack Obama today defended the creation of a Palestinian state. At the end of his first meeting with President of the Palestinian Authority, the North American leader once again reiterated its call on Israel to put an end to building settlements erected in the Palestinian Territories and honor the commitments it made.

Both sides said Obama in the White House have "obligations to the roadmap" - the 2003 international plan to resolve the Israeli-Palestinian conflict. These include "stop the colonization." During the discussion with the new Israeli Prime Minister, Benjamin Netanyahu, the last week, "I was very clear about the need to halt the colonization," Obama also clarified.

The Palestinians in turn must make progress in improving their security forces and the reduction of "inciting" anti-Israel called. "I am a big believer of the solution of two states," said Obama, saying he was "confident" the possibility of progress towards peace between Israelis and Palestinians.

Investigating non-suspect texts

The renewal of the Toural square in the center of Guimarães, will move to the end of the year, but the design is totally different from the planned study presented two years ago. The project challenged by vimaranenses resolve the tunnel road and underground parking.

The car traffic will be maintained throughout the area, but there will be news. It is planned to create a street in the far east of Alameda de S. Damasus, within what is now the garden, and to distribute the traffic from the city center. The remaining garden is enhanced with more plant species, and have a new design, giving an idea of urban forest.

The project, coordinated by Maria Manuel Oliveira, the department of architecture at the University of Minho, provides the return of the fountain of Toural, public source of the sixteenth century passed, about one hundred years, the garden of Caramel. One of the central ideas expressed by the architects is the reuse of existing elements, such as furniture.

Investigating non-suspect texts: a real case

2009-05-28 23:25:00 PÚBLICO, Agências

O Presidente Barack Obama defendeu hoje a criação de um Estado palestino. No fim do seu primeiro encontro com o presidente da Autoridade Palestiniana, o líder norte-americano repetiu uma vez mais o seu apelo a Israel para que ponha fim à construção nos colonatos erguidos dos Territórios Palestinos e honre os compromissos que assumiu.

May 29, 2009 - AFP

US President Barack Obama has renewed pressure on Israel over settlements but rejected a timetable for his peace drive, noting domestic pressures heaped on Israeli Prime Minister Benjamin Netanyahu.

As Mr Obama met Palestinian leader Mahmud Abbas for the first time as president, he called for a halt to settlement building on the occupied West Bank, as his administration sparred with Israel over the sensitive issue.

Investigating non-suspect texts: a real case

As duas partes, afirmou Obama na Casa Branca, têm “obrigações face ao roteiro” – o plano internacional de 2003 para a resolução do conflito israelo-palestiniano. Nestas inclui-se “parar com a colonização”. Durante a discussão com o novo primeiro-ministro israelita, Benjamin Netanyahu, a semana passada, “fui muito claro quanto à necessidade de travar a colonização”, esclareceu ainda Obama.

The US president recalled that last week he had been “very clear” with Mr Netanyahu about the need to “stop settlements” and again stated his desire to see a two-state solution to the Israeli-Palestinian conflict.

“I would like to take this opportunity to affirm to you that we are fully committed to all of our obligations under the roadmap, from the 'A' to the 'Z,'" he said.

but has made little progress since it was drafted in 2003.

Investigating non-suspect texts: a real case

Mahmoud Abbas sublinhou, por seu turno, a urgência de tais progressos, declarando que “**o tempo [é] um factor essencial**” no processo.

O apelo ao fim da colonização na Cisjordânia e em Jerusalém Oriental já tinha sido feito na véspera pela secretária de Estado, Hillary Clinton: “**Nenhuns colonatos, nem exceções para o crescimento natural**”.

while Mr Abbas pledged to live up to all previous peace agreements and warned “**time is of the essence**” for a two-state solution.

Ms Clinton said Mr Obama “wants to see a stop to settlements. **Not some settlements, not outposts, not natural growth exceptions.**”

Investigating non-suspect texts: a real case

E já hoje, antes do encontro entre Abbas e Obama, Israel reagira pela voz do porta-voz do Governo, que explicou que o futuro dos colonatos só será decidido através das negociações com os palestinianos. “Entretanto, temos de permitir que a vida continue normalmente nestas comunidades”, disse Mark Regev. O que isso significa é que mesmo que não sejam construídos novos colonatos, a expansão dos já existentes poderá prosseguir.

“Normal life” will be allowed in settlements in the occupied West Bank, government spokesman Mark Regev said, using a euphemism for continuing construction to accommodate population growth.

He added the fate of settlements “will be determined in final status negotiations between Israel and the Palestinians and in the interim, normal life must be allowed to continue in those communities.”

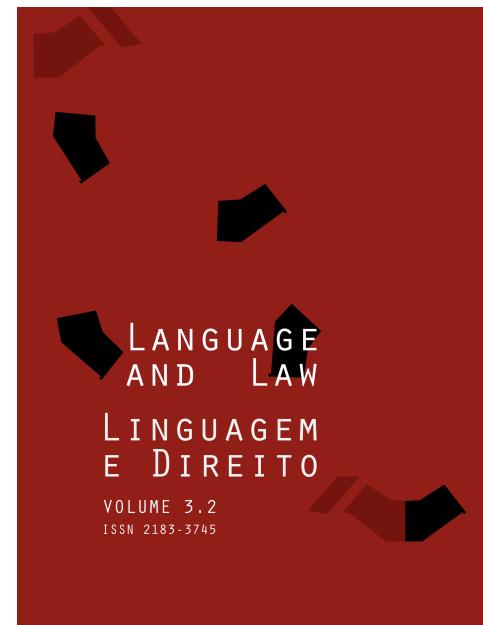
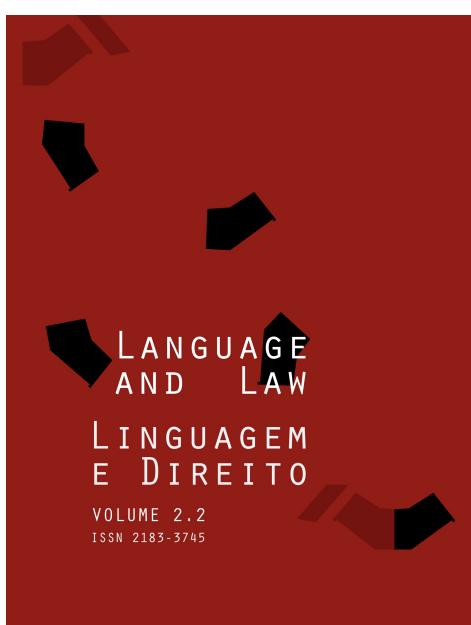
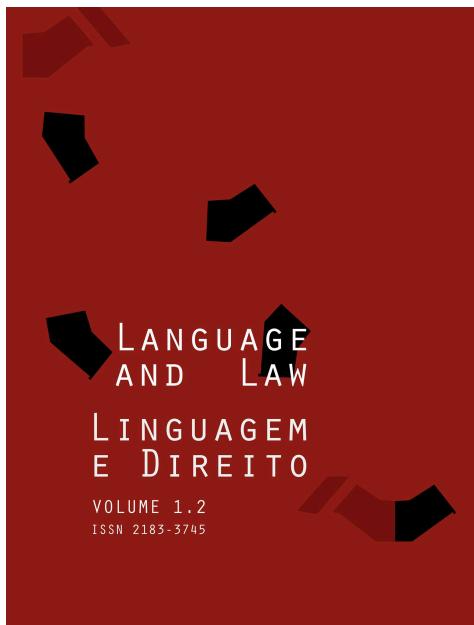
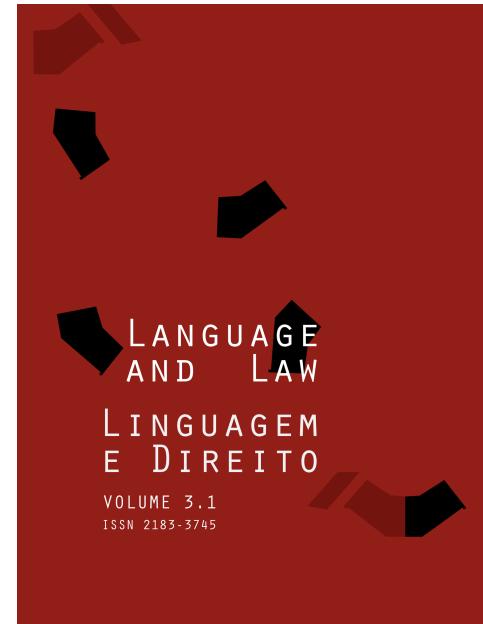
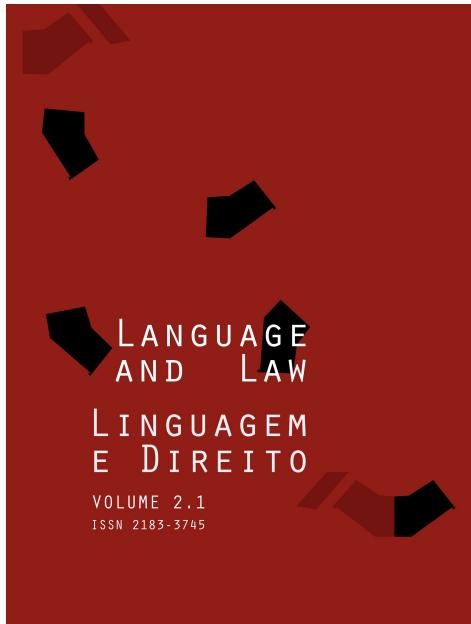
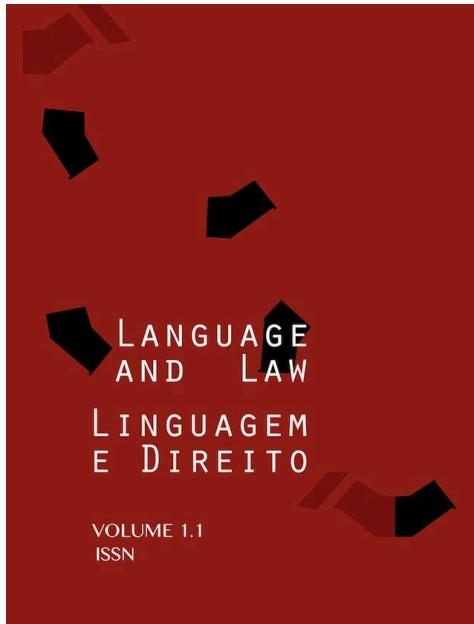
Final (...or not so final...) Remarks

- › Same-language text comparison fails to detect cases where a different language is used to plagiarise.
- › Main challenges: Linguists are competent to compare texts; Two texts in different languages cannot be compared.
- › Proposed solution: back-translating suspect text to suspect source language for comparison.
- › MT handles lexical items and technical terminology very well.
- › MT often fails to handle syntax properly.
- › Same-language comparison based on lexical items, not identical strings.



Final (...or not so final...) Remarks

- › Translingual plagiarism can be detected by a close analysis of (Sousa-Silva, 2013):
 - indices of foreignness.
 - error checking.
 - lexical overlap.
- › Machine translation problem with languages other than English.
- › Other areas where computational linguistics plays a crucial role.
- › Promising future of plagiarism detection:
 - the potential of the focus on lexical items;
 - the use of word lemmas for comparison;
 - the use of the Internet as corpus.

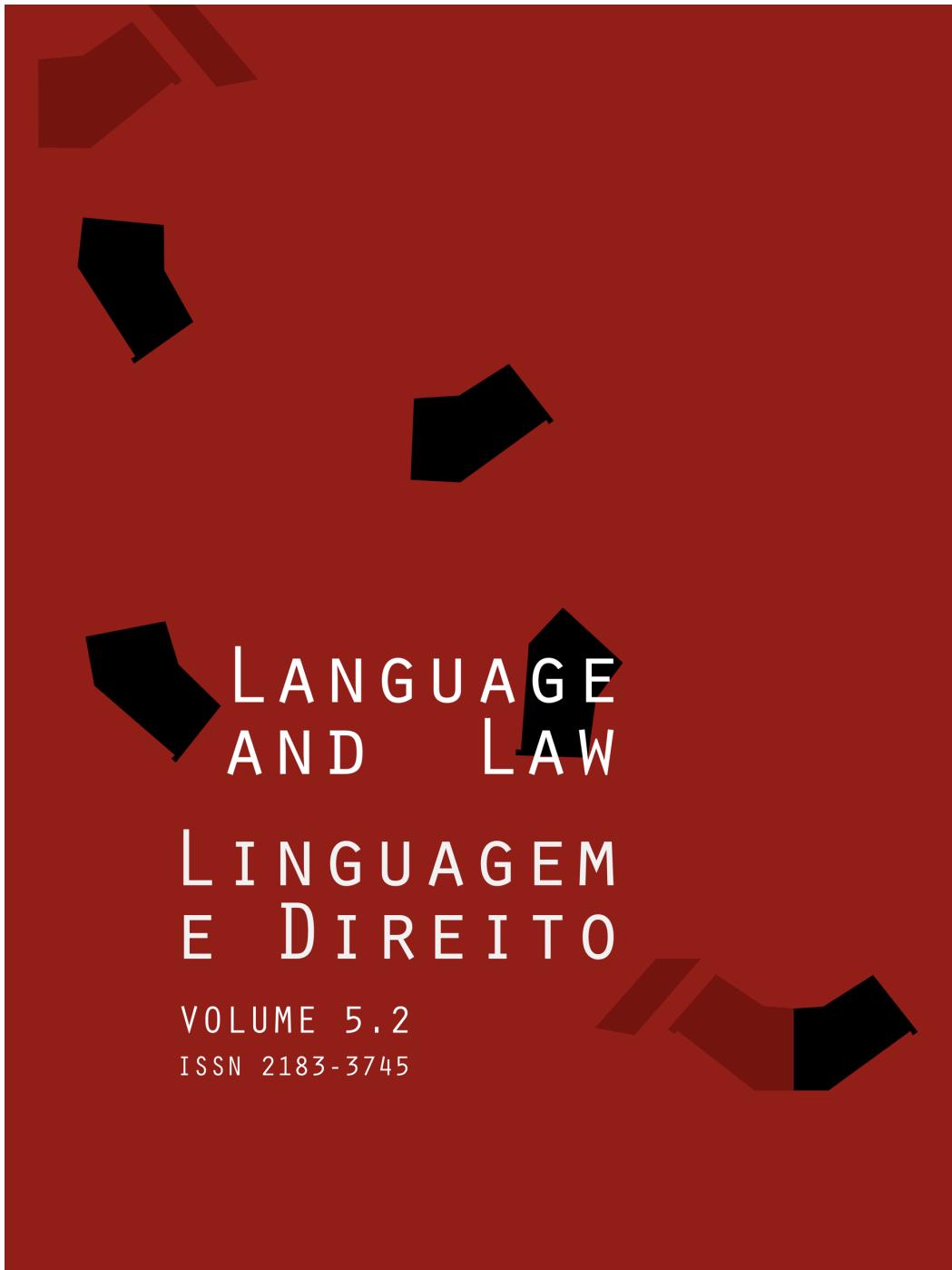


Computational Forensic Linguistics: Special Issue

Vol. 5.2

December 2018

<http://lld.linguisticaforense.pt>



References

- › Bassnett, S. & Lefevere, A. (1998) *Constructing Cultures: Essays on Literary Translation*. Clevedon: Multilingual Matters.
- › Johnson, A. (1997) "Textual kidnapping - a case of plagiarism among three student texts?". *The International Journal of Speech, Language and the Law*, 4(2).
- › Jones, M. (2009) Back-translation: the latest form of plagiarism. *The 4th Asia Pacific Conference on Educational Integrity*. Wollongong: 1-7.
- › Koehn, P. (2010) *Statistical Machine Translation*. Cambridge: Cambridge University Press.
- › Selinker, L. (1972) "Interlanguage". *International Review of Applied Linguistics*, 10.
- › TAUS (2010). MT Post-editing Guidelines, disponível em:
<https://www.taus.net/academy/best-practices/postedit-best-practices/machine-translation-post-editing-guidelines>
- › Turell, M. T. (2008) Plagiarism. In J. Gibbons & M. T. Turell (Eds.), *Dimensions of Forensic Linguistics*. Amsterdam/Philadelphia: John Benjamins. 265-299.
- › Turell, M. T. (2013). Presidential address. In Sousa-Silva, R., Faria, R., Gavaldà, N. & Maia, B. (Eds.), *Bridging the Gap(s) between Language and the Law: Proceedings of the 3rd European Conference of the International Association of Forensic Linguists*. Porto: Faculdade de Letras.

Detecting translilingual plagiarism: A forensic linguistic contribution to computational processing.

This work was partially supported by grant SFRH/BD/47890/2008 and post-doctoral research grant SFRH/BPD/100425/2014, FCT-Fundação para a Ciência e Tecnologia, Portugal, cofinanced by POPH/FSE.