

# Capítulo 6

## Análise e Detecção de Plágio em Contextos Forenses

*Rui Sousa-Silva*

### PLÁGIO: UMA MATÉRIA FORENSE

A ocorrência de plágio tem sido amplamente estudada ao longo do tempo, e sobretudo ao longo das últimas décadas. Embora não seja um fenómeno recente (ver e.g. Sousa-Silva e Abreu 2015 para uma contextualização histórica), a sua ocorrência tem sido despoletada nas últimas décadas, decorrente dos desenvolvimentos tecnológicos e do crescente volume – e massificação – da informação disponível on-line. Essas tecnologias facilitaram a reprodução de informação de modo instantâneo, sem um controlo adequado da sua autoria, dificultando, simultaneamente, a deteção dessa reprodução e a identificação da fonte original. De tal modo que os próprios programas informáticos de deteção de plágio – incluindo o software comercial, que possui uma robustez elevada – dificilmente conseguem identificar a fonte original; antes, ao procurar plágio num determinado texto, é altamente provável que a reprodução (ilícita) de determinado trecho seja mais facilmente atribuída a um texto que reproduziu o texto do original do que à fonte original. Ou seja, é mais provável que a sobreposição textual seja associada a uma fonte secundária do que à fonte primária.

Por isso, o estudo de plágio é frequentemente comparado a uma "caixa de Pandora": ao estudar plágio enquanto problema forense, o linguista tem de ter em consideração aspetos como: a anterioridade de produção – o que, conforme defende

Turell (2007), pode ser complexo em casos de produção contemporânea de textos); os conhecimentos do suspeito plagiador em matéria de redação textual (tipicamente, acadêmico-científica); o tipo de material sujeito a plágio; as estratégias utilizadas para plagiar; o gênero textual; o tipo de plágio; as questões culturais relativamente ao Direito de autor; o conceito de propriedade intelectual; e o possível grau de intencionalidade do suspeito – sem esquecer a possibilidade de duas pessoas diferentes, em momentos e locais diferentes, terem ideias idênticas. Esse foi o caso que descrevi anteriormente (2019) sobre a utilização da metáfora da caixa de Pandora com a análise e deteção de plágio em linguística forense que, descobri posteriormente, já tinha sido previamente utilizada por Sutherland-Smith (2005). O argumento de que duas pessoas em momentos distintos podem ter ideias idênticas pode ser, porém, um argumento falacioso, que é utilizado frequentemente por plagiadores como desculpa para normalizar ou desculpabilizar o ato de plagiar.

Não obstante todos os condicionamentos descritos, o plágio é considerado (quase) universalmente como sendo condenável, tendo adquirido nos últimos anos maior visibilidade devido a casos mediáticos a acontecer um pouco por todo o mundo, não só no contexto académico, mas também em contextos não académicos, como o jornalismo ou a política. De fato, os casos mais mediáticos têm sido os relacionados com políticos. Neste campo, ficaram célebres os casos dos ministros alemães da defesa, zu Guttenberg (em 2011), e da educação, Annette Schavan (em 2013). Conhecido também foi o caso do Primeiro-ministro romeno Victor Ponta (em 2012). Os dois ministros alemães, que foram alvo de denúncias anónimas, viram as suas teses anuladas pelas universidades que lhes conferiram o grau e acabaram por se demitir. O Primeiro-Ministro romeno negou as acusações de plágio e continuou no poder, inclusivamente depois de um júri independente da Universidade de Bucareste ter confirmado as acusações de plágio. Acabou, contudo, por pedir a demissão posteriormente, na sequência de vários casos de fraude em que esteve alegadamente envolvido. Em Portugal, em 2014 o Secretário de Estado do Ensino Básico e Secundário, Dr. João Grancho, demitiu-se após a publicação de várias notícias que reportavam atos de plágio que o político teria cometido sete anos antes, quando presidia à Associação Nacional de Professores de Portugal.

No jornalismo, foram noticiados vários casos paradigmáticos nas últimas décadas. Um desses casos ocorreu em 2011 e envolveu Johann Hari, jornalista do jornal britânico *The Independent*. Hari foi acusado de plágio por utilizar citações em entrevistas sem atribuir o texto aos autores originais. No ano seguinte, Jonah Lehrer,

jornalista da revista Nova-iorquina *New Yorker*, demitiu-se na sequência do escândalo de reciclagem de publicações no blog da revista (para além de ser acusado de outros atos transgressivos, como fabricação de citações). Um dos casos de plágio mais famosos no jornalismo foi, contudo, aquele que envolveu Jayson Blair, em 2003. Este ex-jornalista do jornal americano *The New York Times* tinha sido acusado de vários atos de fraude na sua carreira jornalística, incluindo plágio de material de agências noticiosas e de outros jornais, tais como o *Washington Post*. Na sequência do escândalo, Blair acabou por pedir a demissão.

No entanto, o plágio no jornalismo também existe em países de Língua Portuguesa. Em 2007, uma jornalista do jornal português *Público* foi acusada de ter traduzido textos de outras fontes, incluindo da *Wikipedia* e da *NewScientist*, reutilizando-os nas suas peças jornalísticas, sem atribuição aos autores originais. O caso, então denunciado por um leitor, foi investigado pelo próprio jornal; inicialmente, a jornalista negou as acusações, mas, perante as evidências apresentadas, acabou por pedir desculpa publicamente aos leitores, permanecendo no jornal.

## PLÁGIO, CONVENÇÕES DISCIPLINARES E GÉNEROS TEXTUAIS

As graves consequências de casos de plágio, cometidos prévia ou contemporaneamente, na vida profissional dos supostos plagiadores são evidentes nos casos dos jornalistas e dos políticos referidos, embora, em diversas esferas profissionais, a condenação do plágio pareça estar intimamente ligada às tradições da respectiva disciplina, nas quais a reutilização de texto, seu ou de outrem, é considerada aceitável. Este parece ser o caso de alguns géneros textuais do domínio jurídico, como o género acórdão. Em 2015, foi divulgado publicamente um acórdão do Tribunal da Relação de Lisboa que possuía mais de quatro páginas semelhantes (ou mesmo idênticas) a uma decisão do Tribunal da Relação de Coimbra, datada de 2010, redigida por um outro juiz. Apesar do volume de sobreposição, o juiz desembargador do Tribunal da Relação de Lisboa descartou qualquer possibilidade de plágio, indicando que o género textual o permite. Que seja do conhecimento público, o caso não teve consequências adicionais graves.

Outro género textual do domínio jurídico no qual é comum ocorrer uma

elevada sobreposição textual é na área dos contratos. Um exemplo flagrante de sobreposição textual é aquele que descrevi previamente em detalhe (Sousa-Silva, 2012). Em 2011, vários países europeus (entre os quais, Portugal, Irlanda e Grécia) entraram numa crise económico-financeira profunda e viram-se forçados a pedir ajuda financeira internacional, recorrendo a uma estrutura designada por *Troika*, constituída pela Comissão Europeia, pelo Banco Central Europeu e pelo Fundo Monetário Internacional. Para assegurar essa ajuda financeira, os países intervencionados (como Portugal, Irlanda e Grécia) tiveram de assinar um "memorando de entendimento" (ou seja, um contrato) com a *Troika*. Tendo em consideração que um memorando de entendimento constitui um contrato, seria de esperar alguma sobreposição textual, embora não uma sobreposição textual idêntica ao de um contrato entre, por exemplo, uma operadora de telecomunicações e um cliente (em que as diferenças se limitam, na maioria dos casos, aos dados dos clientes, mantendo-se as condições contratuais; porém, não seria de esperar uma sobreposição elevada no caso do memorando, sobretudo porque se tratava de três países profundamente diferentes, com contextos e infra-estruturas distintas, e com necessidades diversas. No entanto, a análise comparativa da sobreposição textual dos documentos assinados individualmente entre a *Troika*, de um lado, e Portugal, a Irlanda ou a Grécia, do outro, chega a ser superior a 80%.

Do mesmo modo, no jornalismo é frequente argumentar-se que as notícias reportam factos e que, por isso, para serem isentas, objetivas e verdadeiras, as peças noticiosas têm que se limitar a reportar factos, o que limita a capacidade criativa do autor e, por conseguinte, aumenta a probabilidade de sobreposição textual. Não obstante este ser um pressuposto amplamente defendido, a verdade, a objetividade e a isenção noticiosas não são violadas por formas de redação alternativas, antes pelo contrário; como demonstram algumas teorias linguísticas, e como se poderá comprovar através de um simples estudo empírico, a probabilidade de uma mesma pessoa utilizar exatamente a mesma formulação linguística para descrever um mesmo facto em momentos distintos é muito ínfima ou mesmo praticamente nula (Coulthard, 2004); menos provável ainda é a possibilidade de duas pessoas diferentes produzirem texto idêntico, mesmo tratando-se da descrição dos mesmos factos. Em casos de plágio é necessário, portanto, ter em consideração as condicionantes disciplinares e de género textual, mas também os contextos específicos em causa.

Indubitavelmente, as consequências de plágio mais graves refletem-se sobretudo na Academia, onde se considera a prática de atos de plágio inaceitável. Em

2010 uma professora de uma instituição de ensino superior em Portugal foi acusada de ter plagiado na sua tese de Doutoramento, depois de uma denúncia anónima à universidade que conferiu o grau informar que a tese era em grande parte copiada de uma tese defendida anteriormente no Brasil. Após uma investigação da universidade que conferiu o grau, concluiu-se que a tese continha trechos extensivamente copiados da tese brasileira, com pequenas alterações, por exemplo, na grafia isto é, alteração da grafia do Brasil para a grafia de Portugal. O título de Doutoramento foi anulado pela universidade conferente do grau e a docente acabou por ser demitida da instituição de ensino superior onde trabalhava.

É claro, porém, que os casos de plágio são abordados de forma distinta, dependendo de se tratar de contextos académicos ou não académicos — e mesmo entre instâncias do contexto académico. Em contexto não-académico, o plágio é normalmente considerado uma violação de Direito de autor, usurpação de propriedade intelectual ou contrafação. Uma vez que estes têm, habitualmente, implicações económicas, são considerados competência dos tribunais, que determinam eventuais sanções e compensações pela violação de direitos morais à paternidade da obra ou de direitos patrimoniais (Pereira, 2003). Pelo contrário, no contexto académico, o plágio tem sido abordado sobretudo como um ato ilícito praticado por estudantes, constituído por "roubo de palavras", pensamentos, processos, resultados e ideias. Consequentemente, esses casos são vistos como uma violação de princípios éticos, ficando o processo à discrição da Academia e não dos tribunais. Na maioria das vezes, são os professores que gerem os casos, penalizando os estudantes pela prática, embora muitas universidades tenham estabelecido que os casos de plágio devem ser denunciados institucionalmente, sendo a instituição responsável por gerir o caso e aplicar possíveis sanções.

Nestas circunstâncias, não é surpreendente que os casos de plágio académico e não académico sejam perspectivados de forma diferente e que se considere o plágio um problema inerentemente académico. Nos últimos anos, contudo, essa perspectiva tem vindo a mudar, tendo ocorrido casos de plágio académico tratados pelos tribunais. Um desses exemplos é o caso de um autarca português que foi investigado pelo Ministério Público de Portugal por plágio na sua tese de doutoramento<sup>1</sup>.

No entanto, seja em contexto académico, seja em contexto não académico, as

---

<sup>1</sup> Disponível em (<https://www.publico.pt/2019/04/12/sociedade/noticia/autarca-torres-vedras-acusado-plagio-tese-doutoramento-1869028>).

acusações de plágio podem ter implicações graves, pelo que é crucial obter provas concretas que confirmem ou infirmem a existência de plágio, de modo a assegurar uma administração de justiça, não apenas legal, mas também social. Como refere Turell (2013), é função da linguística forense tratar, não só de casos que sejam do interesse do fórum, no sentido restrito do domínio legal, mas também do fórum no seu sentido lato, enquanto sociedade. A análise linguística forense encontra-se idealmente posicionada para desempenhar esta função, na medida em que permite obter pistas linguísticas que indiciam falhas de produção independente de determinado texto, e, por conseguinte, revelam tratar-se de plágio, como também averiguar a direcionalidade (isto é, que texto copiou qual), determinar quais as estratégias utilizadas para plagiar, provar por que motivo não se pode considerar determinado texto como sendo original e auxiliar na determinação do grau de intencionalidade do autor.

Este aspeto é essencial. Com os desenvolvimentos tecnológicos e com a oferta de programas informáticos de auxílio à deteção de plágio, em contextos académicos muitos professores passaram a acreditar cegamente nas estatísticas de reprodução textual apresentadas pelo software após a submissão. Porém, nenhum relatório deve ser considerado cegamente, uma vez que poderão ocorrer diversos tipos de sobreposição textual que não são plágio. É o caso, por exemplo, da citação adequada, em que o autor coloca aspas e a fonte; embora várias ferramentas computacionais atuais já permitam excluir fontes citadas, nem sempre esta função é executada corretamente. O mesmo acontece com citações longas ou, no caso de textos na área das ciências sociais e humanas, anexos e transcrições. Adicionalmente, é comum o software considerar cadeias de palavras relativamente curtas, que, por serem comuns, não estão sujeitas a "propriedade individual", mas que, no seu conjunto, podem contribuir para aumentar a percentagem de sobreposição.

Mais importante ainda é considerar as competências de redação dos suspeitos. Como defende Howard (1995), em casos de plágio académico é necessário distinguir aqueles que são praticados por estudantes experientes, nos anos mais avançados da sua formação, daqueles que são praticados por estudantes em formação inicial, como acontece com estudantes no primeiro ano de licenciatura. Enquanto os primeiros, dada a sua formação, deverão já ter adquirido a competência para escrever academicamente, os estudantes em início de formação ainda estão a aprender como escrever, praticando aquilo a que Howard se refere como "patchwriting". O conceito de "patchwriting" descreve um processo de escrita semelhante a uma manta de

retalhos, em que o/a estudante vai justapondo trechos de texto que, decorrente de uma tentativa fracassada de escrever academicamente, por exemplo recorrendo a parafraseamento, resulta em potencial plágio. Esta prática decorrente de falta de competência, no entanto, é identificável e interpretável linguisticamente, permitindo sustentar a tese de que o/a estudante não teve intenção de plagiar, pois pressupõe que, não obstante a reutilização de trechos de texto (possivelmente sem aspas), a fonte original é fornecida.

Por essa razão, partindo do pressuposto de que reutilizar texto não é sinónimo de plagiar, é essencial distinguir sobreposição textual de plágio.

## DEFINIÇÃO DE PLÁGIO

Uma discussão central em torno da prática de plágio é a própria definição, uma vez que é esta definição que permite averiguar aquilo que constitui ou não plágio, bem como traçar as fronteiras entre plágio e a reprodução lícita de texto. A preocupação com casos de plágio é transversal a praticamente todas as áreas de atividade, e aquela que era, até há poucos anos, uma questão predominantemente académica passou, também, às mais diversas comunidades de prática. Assim, também no domínio profissional passou a prestar-se mais atenção àqueles que, sobretudo ocupando cargos de sucesso, podem alguma vez ter plagiado. Esta abordagem à possível existência de plágio é, por isso, muitas vezes exacerbada por aquilo a que Cohen (1972) designou por "pânico moral", atitude que realça a perversidade do ato de plagiar e defende a sua penalização implacável. Esse "pânico moral" leva os membros de determinada sociedade a estarem atentos ao próximo, em vigilância permanente, em posição de denunciar e de penalizar um crime praticado contra a sociedade; trata-se, de certa forma, de uma atitude centrada naquilo a que Foucault (1977) chamou "vigiar e punir".

Indubitavelmente, a palavra "plágio" tem como origem uma conotação negativa, associada a atos de "roubo", "rapto" ou "violação", entre outras, sendo frequentemente referido como "crime de plágio". Etimologicamente, uma pesquisa em diversos dicionários revela que, no latim, a palavra era utilizada para denominar crimes relacionados com a comercialização ilícita de escravos, ao passo que, no grego, a palavra significa "oblíquo", "trapaceiro", "enganoso". Mais recentemente, a palavra

tem sido utilizada como referência à utilização indevida ou à apropriação de trabalhos alheios como propriedade sua.

Neste contexto, não podemos deixar de colocar algumas questões que constituem desafios éticos contemporâneos, quer em ambiente acadêmico, quer em ambiente não acadêmico: em que consiste o plágio? O que é e o que não é plágio? Poderemos presumir que existe uma definição universal de plágio? Será eticamente aceitável considerar-se como plágio a reutilização textual por determinado estudante, mas desculpar-se a reutilização textual por parte de um profissional – como, por exemplo, um juiz ou um jornalista – ou por uma instituição, como é o caso do Memorandum assinado com a Troika referido acima? Estará a penalização por plágio associada ao volume de reutilização textual? Será o software de deteção a solução para o problema, ou levantará esse software ainda mais desafios éticos? Qual o papel da linguística forense na prevenção e deteção?

Atentando à forma como diferentes especialistas e instituições definem "plágio" facilmente constatamos que não existe um acordo universal relativamente a essa definição, devido à diferente organização concetual nas diferentes regiões geográficas, culturas e áreas disciplinares. Algumas das diferenças concetuais mais notadas são aquelas que dizem respeito, por exemplo, às questões autorais na "tradição ocidental" e na "tradição oriental"; porém, mesmo dentro da designada "tradição ocidental" existem, certamente, diferenças concetuais. Conforme revelou um estudo realizado anteriormente (e descrito em Sousa-Silva, 2013), existem perceções diferentes relativamente ao plágio, à sua gravidade e penalização mesmo entre países europeus, como é o caso de Portugal e do Reino Unido. Dispor de uma definição clara de plágio é, por isso, essencial, sobretudo para aplicação em contextos forenses.

No âmbito da minha investigação, tenho considerado como plágio "a apresentação intencional de ideias, trabalho e texto de outrem sem uma citação adequada, clara e precisa" (Sousa-Silva, 2013, p. 60). Nesse sentido, poderá considerar-se plágio, não só a reutilização, sem citação, de trechos de texto, mas também de trabalhos e ideias, incluindo, por exemplo, referenciais teóricos ou modelos de análise. Igualmente, pode considerar-se plágio quando o suspeito fornece uma fonte no texto, mas essa fonte está errada ou incompleta, não permitindo a sua verificação. Pelo contrário, não deve considerar-se plágio a utilização de um estilo de referência distinto daquele que está estabelecido. Por exemplo, se determinada

instituição estabelecer como norma de referência a ABNT e um estudante utilizar outra norma (e.g. a APA), esse estudante deve ser penalizado pela qualidade do seu trabalho (nomeadamente por não ter seguido as normas fornecidas) e não por plágio, uma vez que a utilização de uma norma divergente da exigida não impossibilita a verificação da integridade do texto.

Outro critério importante na avaliação de casos de plágio é a intenção do suspeito. Diversos autores têm indicado que uma das causas de plágio, principalmente na academia, é a falta de conhecimentos dos estudantes sobre escrita acadêmica (Angèlil-Carter, 2000; Carroll, 2001; Howard, 1995, 1999, 2001; Pecorari, 2008; Howard e Robillard, 2008). Por isso, obter indicações sobre o grau de intenção de determinado autor suspeito de plágio é fundamental, principalmente em contextos forenses. Muito embora, no âmbito forense, não seja função do linguista determinar o grau de intenção do suspeito nem avaliar a sua culpabilidade, a sua perspectiva sobre a forma de produção do texto permite ao linguista forense avaliar a possibilidade de o suspeito ter intenção de plagiar ou, pelo contrário, a possibilidade de esse plágio ser acidental. Esta avaliação da intenção do autor não pode, contudo, confundir-se com a avaliação de intenções realizada, por exemplo, por psicólogos. Desde sempre os linguistas se têm dedicado a estudar a intenção comunicativa do autor, como, por exemplo, na área da pragmática (Austin, 1962; Searle, 1969), da análise de discurso (Coulthard, 1977) ou da linguística aplicada (Coulthard, 2004). Desse modo, não se trata de avaliar a intenção psicológica do autor, mas sim a sua intenção linguística. Analisando o texto, cabe ao linguista forense estudar quais as ocorrências de sobreposição textual e o(s) tipo(s) de sobreposição encontrado(s) e verificar se as fontes são omitidas (parcial ou integralmente) ou, antes, utilizadas de modo incompetente. Em caso de falta de formação em escrita acadêmica, será de esperar a sobreposição ilícita de texto, mas este vem acompanhado de uma indicação, ainda que insuficiente, de que esse texto não é seu; é o caso, por exemplo, de tentativas fracassadas de parafrasear. Em casos de intenção, pelo contrário, espera-se uma manipulação do texto, sem qualquer atribuição às fontes corretas, com o objetivo de ocultar o plágio. Neste sentido, quanto maior for o esforço de alteração do texto, numa tentativa de o afastar do original, sem atribuir a(s) ideia(s) ao(s) seu(s) autor(es), maior será a intenção de plagiar. É claro que a reutilização literal de texto sem atribuição às fontes também representa plágio intencional; porém, uma vez que este é facilmente detetável, será menos grave — ou meramente negligente — do que um texto profundamente manipulado.

No caso do ensino superior, a prática de plágio é, normalmente, considerada mais grave quanto mais elevado for o grau de estudos; inversamente, a sua ocorrência é perspectivada como sendo proporcionalmente menos grave no caso de estudantes de mestrado e, sobretudo, de licenciatura. Este princípio, exposto inicialmente por Howard (1995) e posteriormente reforçado por outros autores (e.g. Howard e Robillard, 2008) assenta no pressuposto de que os estudantes em início de formação (como, por exemplo, de licenciatura) ainda se encontram em processo de aprendizagem de escrita científica; recorrem, por isso, a processos de escrita por imitação (muitas vezes, indevida) das fontes, construindo um texto que mais parece uma espécie de manta de retalhos. Pelo contrário, essa não é admissível a estudantes em ciclos de estudos mais avançados, como é o caso de mestrado e doutoramento, uma vez que se pressupõe que esses estudantes já terão adquirido os conhecimentos necessários a uma utilização adequada e devida das fontes. Não obstante a relativamente menor gravidade aplicável a casos de plágio cometido por estudantes em início de formação, em diversos países (como é o caso do Reino Unido) as consequências para os estudantes são, por vezes, extremamente sérias — até desproporcionais —, implicando em certas situações a sua expulsão da instituição de ensino superior.

Nesse sentido, a linguística forense desempenha um duplo papel na investigação de casos de plágio: contribuir, por um lado, para a deteção de textos produzidos através da prática de plágio (a que designo de "textos plagiadores") e, por outro, analisar linguisticamente o texto suspeito e os possíveis originais (a que designo "textos plagiados"), com o objetivo de identificar os processos (lícitos ou ilícitos) de reutilização de textos previamente disponíveis, as estratégias utilizadas pelo(a) suspeito(a) para plagiar, averiguar o possível grau de intencionalidade do(a) suspeito(a) e, finalmente, ajudar os decisores (sejam eles instituições legais, como os tribunais, ou académicas, como as universidades) a compreender se determinada formulação é comum (sendo de esperar no contexto em causa, e não representando por isso um caso de plágio) ou se a formulação original é tão idiossincrática que a probabilidade de ser proposta por dois autores diferentes seria muito reduzida ou praticamente nula.

É de realçar que a terminologia utilizada, "plagiador" vs. "plagiado", é nova; habitualmente, designa-se por "texto plagiado" aquele que, baseando-se num texto previamente existente, e copiando-o integral ou parcialmente, não é original. Porém, esta opção não é a mais rigorosa; considerando as noções de agente e paciente, aquele

que exerce a ação e aquele sobre o qual a ação é exercida, será mais rigoroso utilizar a designação de "plagiador" para aquele que canibaliza textos previamente existentes, os textos "plagiados".

Vejamos, em seguida, algumas das principais estratégias linguísticas utilizadas para plagiar.

## ESTRATÉGIAS DE PLÁGIO

Independentemente da motivação subjacente ao ato de plagiar, as estratégias de plágio utilizadas são recorrentes, incluindo a cópia literal, a cópia parcial, o parafraseamento e a tradução.

### *CÓPIA LITERAL*

O tipo de plágio mais comum é o plágio literal, em que o texto original é copiado do original, palavra por palavra, sem alterações ou com alterações mínimas, e sem qualquer referência ou atribuição do texto às fontes. Vejamos o seguinte exemplo, retirado do corpus *CORRUPT – Corpus of Reused and Plagiarised Texts*, um corpus pessoal de textos acadêmicos e não acadêmicos reais, que foram investigados e acusados de plagiar textos originais:

Original	Plágio
Esta teoria pode parecer familiar quando se pensa nos esforços que os impressionistas fizeram para captar a luz ou as cores num momento determinado.	Esta teoria pode parecer familiar quando se pensa nos esforços que os impressionistas fizeram para captar a luz ou as cores num momento determinado.

Quadro 1: Plágio Literal

Este tipo de plágio é facilmente detetável, uma vez que, em caso de suspeita de plágio, é suficiente fazer uma pesquisa num motor de busca da Internet comum para obter como resultado textos que utilizam trechos idênticos. Obviamente, nesse caso o linguista forense terá de ter em conta a data de produção dos textos, uma vez que os resultados devolvidos pelo motor de busca podem incluir textos escritos prévia ou posteriormente ao texto suspeito, ou escritos por outros autores ou pelo mesmo autor. Outra forma de detetar facilmente instâncias de plágio consiste em recorrer a software habitualmente utilizado para detetar plágio, como o *Urkund*, o *SafeAssign* ou o *Turnitin*, entre outros. O recurso a software especializado deste tipo apresenta diversas vantagens, como sejam pesquisar entre bases de dados de textos disponíveis on-line e identificar os originais, como também assinalar os trechos de texto que são idênticos. Este sistema de visualização de semelhanças facilita a tarefa do linguista, permitindo-lhe focar nas alterações operadas nos textos. A principal desvantagem desses sistemas é a sua disponibilidade: tendo em conta a sua estrutura e os custos envolvidos, muitos deles estão, normalmente, disponíveis apenas para subscrição institucional, sendo o acesso vedado a utilizadores individuais. Nesse tipo de comparação, em que o linguista forense não possui outros textos para comparar com o texto suspeito, o objetivo da análise é procurar textos idênticos on-line. Porém, sempre que o linguista possui textos relativamente aos quais pode comparar o texto suspeito, essa tarefa pode ser executada localmente, sem necessidade de utilizar ferramentas disponíveis on-line. Nestes casos, a comparação local pode ser realizada recorrendo a software específico, como é o caso do *CopyCatch* ou do *WCOPYFind* (este último, de utilização gratuita), ou mesmo a software genérico de processamento de texto. No caso de programas de processamento de texto, uma comparação de documentos é suficiente para mostrar as semelhanças e as diferenças entre as fontes consideradas originais e o texto suspeito.

### CÓPIA COM ALTERAÇÕES

Apesar de existirem, os casos de cópia literal, em que o texto original é reutilizado sem alterações (e sem atribuição), são raros em situações de plágio intencional, uma vez que são facilmente detetáveis. Mais comum é a cópia parcial, em que o suspeito recorre a uma manipulação do texto, introduzindo alterações que tornam mais difícil a sua detecção.

Em 2017, foi publicada na Internet uma entrada com o título "Como não ser pego no plágio – guia definitivo!", assinado por Yuri Ferreira<sup>1</sup>. Nesse texto, o autor admite que a atenção dedicada à detecção de plágio levou ao desenvolvimento de mais e melhores ferramentas de detecção de plágio, o que dificulta a tarefa de plagiar. Por isso, propõe-se a "dar as melhores dicas de como não ser pego no plágio", "de como não ser acusado de cópia" e de como "tentar encontrar falhas nos programas de detecção de plágio". A primeira dessas "dicas" é "Não seja preguiçoso na hora de ser preguiçoso!": partindo do princípio de que o plágio literal é facilmente detetável, o autor do texto defende que é necessário fazer alterações ao texto, de modo a passar como texto original. Propõe, por isso, alterações ao texto original, nomeadamente a reorganização de palavras, a substituição de palavras por sinónimos, a alternância de textos e a mudança de citações. O software utilizado mais frequentemente em trabalhos de detecção de plágio baseia-se numa pesquisa de trechos de texto no documento suspeito que sejam idênticos a texto publicado anteriormente. Essa procura baseia-se em cadeias de palavras, pelo que, quanto mais longas forem essas cadeias, mas facilmente são identificados esses trechos de texto. Por isso, será lógico pensar-se que reorganizar as palavras, substituir palavras por sinónimos, alternar textos e mudar citações irá interromper essas cadeias de texto mais longas, e, conseqüentemente, dificultar o processo de detecção.

Essas estratégias, contudo, não resistem a uma análise linguística forense. Duas delas são a alternância de textos e a mudança de citações. As duas, mas sobretudo a última, são úteis para tentar enganar o leitor (por exemplo, os professores responsáveis pela avaliação do trabalho), mas não enganam os sistemas automáticos: ao mudar citações, atribuindo a determinado autor um texto que é de outrem, direciona o leitor para uma referência errada, impedindo a confrontação com o original; porém, o software permite verificar a sobreposição do texto com a fonte original. O mesmo acontece com a alternância de textos: imagine-se que o plagiador copia de três fontes – "Texto A", "Texto B" e "Texto C" – e, ao invés de copiar:

---

1 Disponível em <https://educacao.umcomo.com.br/artigo/como-nao-ser-pegono-plagio-guia-definitivo-28288.html>

TEXTO A. TEXTO B. TEXTO C.

na íntegra e por esta ordem, copia uma ou duas frases de cada texto:

FRASE DO TEXTO A. FRASE DO TEXTO C. FRASE DO TEXTO A. FRASE DO TEXTO B. FRASE DO TEXTO C. FRASE DO TEXTO B. FRASE DO TEXTO A. FRASE DO TEXTO C. FRASE DO TEXTO B. FRASE DO TEXTO A.

Apesar desta alteração, muito dificilmente essa estratégia funciona. Primeiramente, ainda que o plagiador utilizasse apenas frases ou trechos de frases dos textos originais, esses trechos são facilmente detetáveis pelo software. Em segundo lugar, essa alteração exigiria um trabalho significativo por parte do plagiador, uma vez que o texto resultante da alternância de trechos teria de fazer sentido, o que exige um trabalho cuidado para assegurar a coesão e a coerência dos textos. Ora, conseguir articular referências deste modo, desde que com a devida citação, é o objetivo da escrita acadêmica; se considerarmos que o principal motivo pelo qual esse objetivo nem sempre é cumprido é a falta de tempo e a preguiça do plagiador, então poderemos pressupor que essa estratégia de plágio é extremamente rara, senão inexistente. Por outras palavras, se o plagiador se der ao trabalho de produzir um novo texto, integrando várias referências – como se espera de um trabalho acadêmico – então irá citar devidamente e colher os créditos do seu trabalho, em vez de plagiar.

Estratégias de plágio distintas são a reorganização de palavras e a substituição de palavras por sinónimos e o parafraseamento. A substituição de palavras por sinónimos é uma das estratégias mais utilizadas para ocultar a reutilização do texto de outrem, uma vez que permite fazer alterações formais ao texto original sem um grande esforço. Esta estratégia consiste em substituir uma palavra ou combinação de palavras por palavras de significado idêntico ou semelhante. Embora essas palavras mantenham algum tipo de relação semântica com o texto original (como sejam relações de sinonímia, hiponímia ou hiperonímia), também podem ser de campos semânticos diferentes, especialmente quando o plagiador procura assegurar a coerência com o mundo extra-textual. Vejamos o seguinte exemplo:

Original	Plágio
Ao fazer a mesma pergunta a diferentes pessoas, podemos conhecer opiniões variadas sobre um determinado tema. Será muito interessante confrontar o que pensam os alunos, <b>os professores e os pais</b> sobre as diversas questões da actualidade educativa.	Ao fazer a mesma pergunta a diferentes pessoas, podemos conhecer opiniões variadas sobre um determinado tema. Será muito interessante confrontar o que pensam os alunos, <b>os profissionais e o público em geral</b> sobre as diversas questões da actualidade cultural.

Quadro 2: Substituição de palavras.

Neste texto, por exemplo, "os professores e os pais" é substituído por "os profissionais e o público em geral", e "educativa" é substituído por "cultural". Este é um cenário comum em casos de plágio académico, uma vez que as alterações são pontuais, tendo como objetivo adaptar o texto plagiador ao contexto aplicável – a criação de um jornal escolar –, exigindo, por isso, ao plagiador um esforço mínimo de substituição de palavras. As alterações podem, contudo, ser mais sofisticadas, como mostra o exemplo seguinte:

Original	Plágio
É absolutamente essencial definir quem se pretende que sejam os principais leitores do jornal. <b>Ao escolher os jovens em geral, os alunos do estabelecimento de ensino, a comunidade educativa ou os habitantes da localidade como o público a quem o jornal prioritariamente se vai destinar, está a ser feita uma</b> opção que condicionará a forma e o conteúdo da publicação.	É absolutamente essencial definir quem se pretende que sejam os principais leitores do jornal. Esta opção irá condicionar a forma e o conteúdo da publicação.

Quadro 3: Substituição de palavras.

Nesse caso, o trecho em negrita é eliminado e o texto plagiador transforma uma oração subordinada, "opção que condicionará a forma e o conteúdo da

publicação" (no original), numa nova frase: "Esta opção irá condicionar a forma e o conteúdo da publicação.". Por isso, altera a forma verbal do verbo principal no futuro, "condicionará", colocando-o no infinitivo, "condicionar", e adicionando um verbo auxiliar, "ir", conjugado no futuro, "irá". Adiciona, também, um novo pronome demonstrativo, ("Esta"), que tinha sido omitido no original, e o pronome relativo "que" é eliminado. Estas alterações têm como resultado uma nova formulação, que, apesar de semanticamente idêntica ao original, é diferente do ponto de vista morfosintático, podendo contornar os sistemas de deteção automática.

A reorganização de palavras utiliza-se para descrever as operações linguísticas através das quais as palavras do texto original são utilizadas no texto plagiador, mas numa ordem diferente. Embora esta estratégia não seja tão utilizada como a substituição de palavras, o seu uso é frequente, como mostra o seguinte exemplo:

Original	Plágio
Adaptando ao fotojornalismo uma sistematização das funções da linguagem no discurso informativo sustentada por Jesus <b>Gonzalez</b> Requena(41), poderíamos (...)	<b>Para Jesus Requena</b> , adaptando ao fotojornalismo uma sistematização das funções da linguagem no discurso informativo poderíamos (...).

Quadro 4: Reorganização de palavras.

Neste caso, o texto plagiador edita o nome do autor ("González" é eliminado), a vírgula e o número da nota são eliminados, e a oração "sustentada por Jesús González Requena" é alterada para "Para Jesús Requena". Resultante destas alterações, o texto plagiador retém uma sequência de 11 palavras do original, cinco das quais são itens gramaticais, de uma cadeia com um total de 25 palavras no original. Este processo pode dificultar a deteção de plágio através de sistemas automáticos, uma vez que, se as cadeias de palavras forem interrompidas de tal modo que o número de palavras idênticas é muito reduzido, a probabilidade de os sistemas assinalarem esses trechos como potenciais casos de plágio é muito reduzida. Contudo, a utilização de algum léxico – como, por exemplo, terminologia específica – é incontornável, pelo que um método que foque na utilização de itens lexicais,

como aquele que é proposto por Johnson (1997), permitirá identificar essa reutilização.

Nesse estudo, Johnson comparou textos académicos de vários estudantes suspeitos de plágio e verificou que, enquanto dois estudantes, em particular, apresentavam cadeias de palavras idênticas, um terceiro estudante apresentava uma menor sobreposição de cadeias de palavras, resultante de uma reorganização de palavras mais aprofundada, não obstante o facto de os argumentos apresentados serem idênticos. Uma vez que as alterações sintáticas, nomeadamente na ordem das palavras, conduzem inevitavelmente à alteração das palavras gramaticais utilizadas, a autora focou a sua análise apenas nas palavras lexicais, descartando as palavras funcionais. Produzindo listas de palavras lexicais, comparou o léxico usado pelos autores suspeitos, independentemente da ordem em que esse léxico foi utilizado, de modo a comparar, quer o léxico comum, quer o léxico mais raro – aquele que ocorre apenas uma vez em cada texto – utilizado pelos autores suspeitos. Johnson concluiu que, apesar de o tema do texto ser o mesmo e de se esperar alguma sobreposição lexical entre os autores, os autores suspeitos apresentavam percentagens de sobreposição muito superiores a outros estudantes (insuspeitos), que integraram um grupo de controlo.

O método proposto por Johnson (1997), e mais tarde descrito em Coulthard e Johnson (2007), permite ao linguista forense concentrar-se no vocabulário utilizado, independentemente da ordem em que esse vocabulário ocorre, tendo sido aplicado em casos de linguística forense, como descreve Turell (2004).

### *PLÁGIO DE IDEIAS*

A reorganização do texto é frequentemente confundida com o parafraseamento. Em textos académicos, é comum e, inclusivamente, desejável recorrer-se à paráfrase como forma de integrar a voz autoral de outros no nosso próprio texto. Contudo, a paráfrase, que consiste em descrever a mesma ideia por outras palavras, é um mecanismo complexo; de facto, uma paráfrase competente exige ao seu autor um domínio, quer da língua, quer do tema do texto, por forma a explicitar a ideia de determinado autor sem reutilizar as mesmas palavras e sem deturpar o sentido do texto original. Por essa razão, as tentativas realizadas por estudantes para parafrasear são, não raras vezes, fracassadas, constituindo mais uma

substituição de palavras e reorganização da sua ordem do que uma paráfrase, o que pode levar a que tentativas incompetentes de parafrasear resultem em acusações de plágio.

Esse é um desafio constante para o/a linguista forense, que, de modo a assegurar conclusões justas, deverá ter em conta: (1) o contexto de produção do texto (incluindo o ano de escolaridade do autor suspeito); (2) os mecanismos utilizados na produção da paráfrase; (3) a inclusão, ou não, da citação dos autores originais. Considerar este último elemento é essencial, pois pode permitir fazer a distinção entre falta de conhecimentos de escrita académica e violação de integridade académica (plágio). No caso da primeira, o autor suspeito utilizaria grande parte do vocabulário e da sintaxe do texto original, mas citaria o seu autor; na segunda, o autor suspeito faria algumas alterações ao texto plagiado, mas omitiria o nome do autor original.

A análise e a deteção de plágio em paráfrases realizadas de forma competente, mas sem citação — ou com uma citação errada — do autor original são significativamente mais complexas. Em primeiro lugar, neste tipo de paráfrases, a probabilidade de o autor suspeito reutilizar trechos de texto do original é ínfima ou praticamente nula, o que significa que muito dificilmente esses trechos serão assinalados por software de deteção de plágio. Em segundo lugar, é pouco provável que o leitor seja intuitivamente levado a pensar que esse texto constitui plágio, exceto nos casos em que esse leitor conheça muito bem a bibliografia na área respetiva e consiga identificar a fonte correta e, conseqüentemente, identificar o texto suspeito como sendo plágio — não textual, uma vez que o texto suspeito não utiliza o texto original, mas sim plágio de ideias, considerando que o texto utiliza como suas as ideias de outrem.

Em muitos casos de plágio de ideias, o/a linguista forense consegue encontrar no texto elementos que lhe permitem assinalar determinado texto como suspeito, como, por exemplo, elementos de coesão e de coerência. De facto, embora se espere de um texto académico que este possua uma estrutura lógica, com mecanismos de coesão e de coerência adequados, é frequente estes dispositivos não serem utilizados de forma adequada em casos de plágio, como mostra o exemplo que se segue:

Original	Plágio
Pues bien, hace años, debajo de mi casa, había en una pared una frase en spray que decía "Putá Burgos" y eso no era un graffiti, era una pintada.	Pois bem, se numa parede há uma frase em spray que diz "Putá Espanha", isso não é um graffiti, é uma pintada.

Quadro 5: Plágio de ideias.

Neste exemplo, os autores suspeitos fizeram uma tradução literal do original, em espanhol, para português, fazendo apenas uma alteração hiperonímica (substituindo "Burgos" - cidade espanhola - por "Espanha" - o nome do país), procurando assim reduzir o grau de especificidade do texto e promover a sua coerência. Porém, a utilização do substantivo "pintada" não é coerente em língua portuguesa, o que, juntamente com outros aspetos, permite levantar suspeitas relativamente à originalidade do texto.

#### *PLÁGIO TRANSLINGUE*

Uma estratégia utilizada com frequência em línguas que não o inglês é o plágio translingue (Sousa-Silva, 2013), em que determinado autor traduz um texto original de outra língua, omitindo a referência ao texto original e utilizando, por conseguinte, a tradução como se fosse um texto original, de sua autoria. Como refere o autor da entrada do blog "Como não ser pego no plágio: guia definitivo!":

[u]ma boa dica para quem deseja plagiar um trabalho e não deseja ser pego na hora de cometer sua irregularidade é buscar por artigos que não estejam na nossa língua. A tradução de uma monografia ou artigo científico de uma outra língua é uma ótima pedida. Especialmente quando falamos de línguas que não são o inglês. A chance desses plágios serem pegos é quase zero, inclusive não existem programas capazes de detectar traduções com facilidade, além do acesso a publicações que não sejam em inglês bastante pequeno pela nossa academia.

Embora algum software de deteção de plágio anuncie a disponibilização, em versão experimental, da deteção de plágio noutras línguas, a sua deteção automática não tem apresentado resultados fidedignos. A solução para este problema reside na

deteção e análise manual, aplicando métodos e técnicas de linguística forense como aquele que foi proposto anteriormente (Sousa-Silva, 2013, 2014).

O método proposto assenta no pressuposto de que, por motivos de eficiência de tempo, o plagiador recorre normalmente a motores de tradução automática, que, atualmente, permitem obter uma tradução imediata de qualidade aceitável, à qual é depois aplicado um processo de edição e revisão na língua de chegada, de modo a corrigir potenciais erros gramaticais resultantes do processo de tradução automática.

A tradução automática possui, contudo, alguns fatores condicionais, o mais importante dos quais é a literalidade da tradução. Graças à evolução tecnológica dos últimos anos, os motores de tradução automática disponíveis gratuitamente on-line permitem traduzir razoavelmente léxico e terminologia, embora seja comum apresentarem alguns problemas gramaticais. Este é o caso, sobretudo, de pares de línguas como Inglês > Português: uma vez que, contrariamente ao inglês, o português é uma língua com flexão de género, é comum as traduções automáticas de inglês para português apresentarem problemas nesta matéria – que, quando não corrigidos por processos de edição, podem indiciar casos de plágio translingue.

Outro elemento que pode indiciar o recurso ao plágio translingue é a sintaxe: como a tradução automática produz habitualmente traduções literais, a obtenção, em língua portuguesa, de frases com uma sintaxe próxima do inglês pode, mais uma vez, indiciar que o seu autor recorreu à tradução de um texto em inglês para português.

O método proposto anteriormente para analisar textos suspeitos de plágio translingue é bastante simples, uma vez que, como referem Coulthard e Johnson (2007), as ferramentas que ajudam a plagiar também ajudam a detetar instâncias de plágio. Por isso, para iniciar o processo de verificação da existência de plágio translingue num texto suspeito é suficiente reverter o processo utilizado potencialmente pelo plagiador. No caso de um texto suspeito de ter plagiado um original em inglês, o método a utilizar seria o seguinte:

1. Recolher o texto (suspeito), escrito na Língua A (por exemplo, em Português), com sintaxe idêntica à língua suspeita do original (neste caso, inglês).
2. Traduzir o texto da Língua A (por exemplo, português) para a Língua B (neste caso, inglês) utilizando um motor de tradução automática (como, por

exemplo, o Google Translate, o Bing, etc.).

3. Procurar no texto traduzido formas linguísticas distintas das formas padrão dessa língua, que funcionam como idiosincrasias desse texto.

4. Seleccionar algumas frases do texto traduzido que sejam possivelmente idiosincráticas.

5. Focar nos itens lexicais, deixando de parte palavras gramaticais.

6. Usar essas palavras lexicais como palavras-chave para fazer a pesquisa na Internet; uma vez que, conforme referido anteriormente, a tradução automática traduz razoavelmente o léxico, mas falha nas palavras funcionais, a utilização de palavras lexicais como palavras-chave irá evitar a devolução, pelo sistema, de resultados com nível elevado de ruído.

7. Procurar expressões traduzidas / suspeitas na Internet.

Os exemplos que se seguem, que fazem parte do corpus CorRUPT, ilustram a aplicação deste método. Na primeira linha (PT), encontra-se o texto suspeito, publicado em português; a segunda linha (PT–EN) apresenta a tradução automática do texto suspeito para inglês; a terceira linha (EN) apresenta o original em inglês, encontrado on-line. O texto sobreposto (idêntico) entre a tradução automática e o original encontra-se assinalado a negrita. Esse método, cuja sustentação teórica e demonstração empírica se encontram detalhadamente descritas em Sousa-Silva (2014), tem sido utilizado com êxito em diversas análises forenses de plágio, quer em contexto académico, quer em contexto não académico (e incluindo o domínio judicial).

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

Quadro 6: Exemplo de plágio translingue.

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

Quadro 7: Exemplo de plágio translingue.

## CONCLUSÃO

A detecção e análise linguística forense de plágio constituem uma das áreas com menor investigação em linguística forense, sobretudo quando comparadas com os estudos em análise de autoria. Uma das razões dessa preferência por outras áreas, em detrimento da análise e detecção de plágio, é a perspectiva (enviesada) de que se trata de uma área facilmente solucionada por software de detecção e que, por conseguinte, dispensa investigação mais aprofundada na área. Esta perspectiva não poderia, contudo, estar mais longe da verdade. Com as evoluções tecnológicas dos últimos anos, têm surgido novas estratégias de plágio para as quais o software nem sempre tem resposta. Por outro lado, a proliferação de recursos bibliográficos on-line e a multiplicação de referências citando as mesmas fontes tem influenciado negativamente a eficácia dos sistemas de detecção de plágio. Esse impacto é observável, por exemplo, na identificação das fontes originais, ao comparar um trabalho académico com as bases de dados acedidas on-line pelo software: enquanto, há alguns anos, os resultados da pesquisa devolviam as fontes primárias — i.e., as fontes originais —, atualmente existe uma tendência para devolver como resultados fontes secundárias, isto é, fontes que citam as fontes primárias. Isso levanta, desde logo, um problema sério: se uma fonte secundária for devolvida como sendo original, a comparação pode basear-se em resultados errados.

Além disso, um dos maiores desafios do plágio não é a sua detecção, mas sim a sua análise: mais importante do que encontrar semelhanças e sobreposição entre dois e mais textos é saber analisar linguisticamente os dados, averiguando se essa

sobreposição é justificável, tendo em conta a natureza comum do texto, ou, pelo contrário, se o texto é de tal modo idiossincrático que a sua reprodução constitui indiscutivelmente um caso de ilicitude. Na Justiça, onde os casos de plágio julgados são cada vez mais frequentes, não é suficiente saber qual a percentagem de sobreposição textual; antes, é crucial fornecer uma explicação linguística dos mecanismos utilizados, bem como uma justificação detalhada da possível violação de princípios éticos e/ou legais. Essa análise não pode ser realizada por qualquer profissional, e muito menos por uma qualquer ferramenta informática; antes, requer uma avaliação cuidada e devidamente fundamentada de um/a linguista forense.

Como referi em Sousa-Silva (2019), a investigação na área de plágio é tão complexa que se assemelha a uma "caixa de Pandora", o que a torna uma área de investigação privilegiada para todos aqueles que tenham interesse na área da Linguística Forense.

## AGRADECIMENTOS

Este trabalho foi parcialmente apoiado pela bolsa de doutoramento SFRH/BD/47890/2008 e pela bolsa de pós-doutoramento SFRH/BPD/100425/2014, FCT-Fundação para a Ciência e Tecnologia, Portugal, cofinanciadas pelo POPH/FSE.