

**Relatório de análise linguística forense
Comparação de autoria de mensagens online**

| | |
|-----------------------|---|
| Data: | 30 de julho de 2015 |
| Especialidade: | Linguística forense e análise de autoria |
| A pedido de: | Gabinete Cibercrime. |
| Assunto: | <p>Comparação de mensagens do Facebook assinadas pelo autor [X] com mensagens publicadas online assinadas por [X], para verificar se são do mesmo autor.</p> <p>Por motivos de confidencialidade, na presente versão do relatório foram omitidos os apêndices e as referências identificativas.</p> |

Rui Sousa Silva
Centro de Linguística, Faculdade de Letras
Universidade do Porto
LFLAB.Linguística Forense
r.sousa-silva@forensis.pt

Índice

| Secção | Descrição | Página |
|---------------|--|---------------|
| 1. | Introdução | 3 |
| 1.1. | Contexto da análise | 3 |
| 1.2. | Preparação para análise | 4 |
| 1.3. | Autor da investigação | 5 |
| 1.4 | Resumo das conclusões do relatório | 6 |
| 2. | Investigação linguística | 6 |
| 2.1. | Documentos | 6 |
| 2.2. | Análise linguística forense de autoria | 7 |
| 2.3. | Análise | 8 |
| 3. | Opinião pessoal | 10 |
| 4. | Conclusões | 11 |
| | Referências | 12 |

Relatório

1. Introdução

1.1 Contexto da análise

O processo investigado neste relatório pericial incide sobre um conjunto de textos com os quais o Gabinete do Cibercrime contactou no âmbito de uma investigação de crimes de usurpação de direitos de autor (descrita como sendo vulgarmente designada pirataria e partilha gratuita de conteúdos). No contexto desta investigação, foi identificada a existência de um determinado website onde eram divulgados conteúdos protegidos por direito de autor (como filmes, séries, revistas e livros) sem o consentimento dos detentores dos respetivos direitos de autor, com intuito lucrativo. Paralelamente, foi detetada uma conta na rede social Facebook, com a mesma designação, onde era anunciada a disponibilização de novos conteúdos no website principal, o "site mãe". Não estando os verdadeiros autores responsáveis pelas publicações no website e na conta do Facebook identificados como sendo os mesmos, a questão central da investigação consiste em determinar se esta conta na rede social é gerida pelo mesmo indivíduo que administra o website ou por terceiros.

Recorrendo a tecnologia de identificação de comunicações informáticas, é, normalmente, possível identificar o endereço IP do equipamento utilizado para fazer as publicações suspeitas. No entanto, este método nem sempre é fiável, na medida em que: (a) o endereço IP, na melhor das hipóteses, permite identificar apenas o equipamento informático utilizado, que nem sempre está associado a um só utilizador; e (b) o recurso a mecanismos de obfuscação de endereços IP dificulta a identificação dos utilizadores reais. Pelo contrário, existindo material linguístico em volume suficiente, uma análise de natureza linguística permite confirmar/infirmar a suposição de que determinado suspeito é o autor dos textos-problema, cuja autoria é questionada.

Uma vez que a divulgação dos materiais protegidos por direito de autor é acompanhada por texto, quer no "website-mãe", quer na rede social Facebook, foi pedida pelo Gabinete Cibercrime uma análise linguística forense, cujo objetivo consiste em efetuar uma comparação dos marcadores linguísticos de autoria presentes nos dois conjuntos de textos no sentido de determinar se os textos-problema que compõem a Amostra B foram escritos por um ou vários autores e, em segundo lugar, se terão sido escritos pelo autor dos textos que compõem a Amostra A (ou seja, para determinar se os textos das duas amostras terão sido escritos pelo(s) mesmo(s) autor(es) ou por autores distintos), confirmando – ou, pelo contrário, infirmando assim – a suposição de que a autoria dos textos nos dois meios é a mesma.

Esta comparação de autoria foi efetuada com base em dois conjuntos de amostras. A Amostra A é composta por um texto elaborado pelo gestor do website, que descreve regras de funcionamento do mesmo, bem como por diversos comentários feitos pelo gestor em resposta a questões de vários utilizadores. A Amostra B é constituída por um conjunto de textos publicados numa conta da rede social Facebook, através da qual o utilizador, para além de anunciar a disponibilização de novos conteúdos,

interagia com os seus seguidores.

Esta análise seguiu a metodologia adotada internacionalmente em casos de disputa de autoria em contextos forenses, e que consiste, em primeiro lugar, em identificar elementos linguísticos idiossincráticos (e, portanto, distintivos face à norma linguística) utilizados consistentemente nos textos que compõem a Amostra A, e que, por conseguinte, podem ser considerados marcadores linguísticos. O mesmo tipo de análise é, subsequentemente, aplicado aos textos que compõem a Amostra B. Segue-se, então, a comparação dos marcadores presentes nas duas amostras, no sentido de determinar se os marcadores linguísticos são utilizados de forma distinta ou idêntica nas duas amostras. Uma utilização distinta de marcadores linguísticos indica uma probabilidade elevada de os textos de cada uma das amostras serem de autores diferentes; pelo contrário, uma utilização de marcadores linguísticos idênticos indica uma probabilidade elevada de os dois conjuntos de textos pertencerem ao(s) mesmo(s) autor(es).

Em casos de análise de autoria é comum recorrer-se à análise de textos adicionais, seja de textos do autor suspeito (sobretudo nos casos em que o volume de texto disponível para análise é escasso), seja de textos de outros possíveis autores (nomeadamente para verificar se os textos suspeitos poderão ter sido produzidos por – ou com o contributo de – terceiros). A comparação de autoria realizada no âmbito desta perícia não incidiu sobre quaisquer textos adicionais, nem do(s) autor(es) suspeito(s), nem de outros autores (como, por exemplo, textos dos utilizadores que interagiram com o(s) suspeito(s) através do website e da rede social Facebook).

No caso dos textos que compõem a Amostra B, não foram consideradas as publicações que anunciam apenas nomes de conteúdos (filmes, séries, etc.), nem publicações compostas por textos reutilizados de outras fontes, uma vez que se trata de texto cuja autoria é de outrem e, por conseguinte, não deve ser considerado para não influenciar o resultado da análise.

A presente análise foi realizada no âmbito do projeto de pós-doutoramento *Laboratório Linguístico Cibercrime: linguística forense (computacional) no combate à cibercriminalidade*, financiado pela FCT – Fundação para a Ciência e Tecnologia, e desenvolvido em colaboração com o Gabinete Cibercrime.

1.2 Preparação para análise

Em geral, a análise de autoria no âmbito da linguística forense consiste em analisar um texto ou conjunto de textos cuja autoria é conhecida no sentido de identificar, em primeiro lugar, elementos idiossincráticos no estilo de escrita do respetivo autor, ou seja, marcadores linguísticos próprios desse autor que, em conjunto, constituem o seu estilo idioletoal. Estes marcadores linguísticos são constituídos por elementos de variação dentro de uma determinada norma (escolhas entre formas linguísticas gramaticais ou aceitáveis (corretas) e desvios decorrentes das opções de seleção e co-seleção linguística) e por desvios face à norma (escolhas que são agramaticais ou inaceitáveis (incorretas), desvios decorrentes de gralhas e desvios decorrentes de erros). Com base nestes elementos, procura-se determinar se o autor dos textos possui um estilo

consistente (isto é, os marcadores linguísticos são utilizados de forma recorrente e não pontual) e distintivo (os marcadores linguísticos formam, no seu conjunto, um grupo idiossincrático) face à norma linguística e aos demais falantes. O mesmo processo é aplicado, em segundo lugar, aos textos suspeitos, no sentido de averiguar se os marcadores linguísticos são utilizados consistentemente nos textos que compõem o conjunto (neste caso, os textos que compõem a Amostra B) ou se existem inconsistências indicativas de interferência e intervenção de um terceiro autor, tendo em conta o pressuposto de que as diversas mensagens podem ser de autoria de mais do que uma pessoa. Os resultados da análise das duas amostras são, finalmente, comparados com o objetivo de identificar elementos de semelhança entre as duas amostras no sentido de confirmar ou infirmar, em primeiro lugar, se os textos suspeitos que compõem a Amostra B foram escritos por um ou vários autores e, em segundo lugar, se terão sido escritos pelo autor dos textos que compõem a Amostra A. As conclusões da análise baseiam-se nesta comparação.

A análise linguística forense de autoria deverá considerar sempre o género textual dos textos analisados, de modo a evitar enviesamentos decorrentes de fatores inerentes ao género do texto e não ao estilo de escrita do autor. No caso específico das Amostras A e B, este problema não se coloca, uma vez que os dois conjuntos de textos, tendo sido publicados online e sendo destinados a um mesmo público-alvo, pertencem ao mesmo género textual.

1.3 Autor da investigação

Rui Sousa Silva é Doutor em Linguística Aplicada, com especialização em Linguística Forense, pela Aston University, em Birmingham, Reino Unido, onde colaborou com o CFL – Centre for Forensic Linguistics. É atualmente pós-doutorando em linguística forense financiado pela FCT – Fundação para a Ciência e Tecnologia e Professor Auxiliar (Convidado) da Faculdade de Letras da Universidade do Porto. É, também, membro permanente do CLUP – Centro de Linguística da Universidade do Porto, onde desenvolve a sua investigação e de cuja Comissão Diretiva e Comissão Científica faz parte. No âmbito da sua carreira de investigação científica, desde 2008 tem vindo a realizar trabalhos práticos, científicos e publicações na área da linguística forense, tendo as suas perícias linguísticas contribuído para a investigação e decisão de casos forenses. Há vários anos que apresenta a sua investigação em congressos nacionais e internacionais, sob avaliação dos seus pares. É editor, juntamente com o Professor Doutor Malcolm Coulthard, da revista científica *Language and Law/Linguagem e Direito*, uma revista bilingue internacional dedicada à Linguística Forense. É, também, membro da associação internacional de linguistas forenses, a IAFL – *International Association of Forensic Linguists*, membro da APL – Associação Portuguesa de Linguística, membro fundador da recém-criada ALIDI – Associação de Linguagem e Direito dos Países de Língua Portuguesa e Presidente do Colégio de Linguística Forense da Associação Portuguesa de Ciências Forenses.

O parecer constante deste relatório reflete na íntegra as minhas conclusões, e baseia-se nos resultados da análise linguística forense obtidos

através da comparação dos marcadores de autoria dos textos suspeitos com os textos cuja autoria é conhecida.

1.4 Resumo das conclusões do relatório

Opinião relativa à autoria do conjunto de textos suspeitos

Considerando a comparação entre os textos que compõem os dois conjuntos de textos (Amostra A e Amostra B):

1. O conjunto de textos que compõem a Amostra A é estilisticamente consistente com o conjunto de textos suspeitos que compõem a Amostra B;
2. Os dois conjuntos de textos partilham 26 marcadores linguísticos distintivos, ou seja, partilham as mesmas características linguísticas altamente distintivas;
3. Com base na prova linguística, e tendo em consideração o conjunto de textos fornecido para análise, o autor dos textos da Amostra A poderá ser considerado o potencial autor dos textos que integram a Amostra B;
4. A prova linguística, isoladamente, não permite identificar o autor dos textos da Amostra A como único possível autor dos textos que compõem a Amostra B; porém, a prova linguística permitiria identificar o autor mais provável entre um conjunto de possíveis autores.
5. Não existe qualquer prova estilística que permita excluir a possibilidade de os textos que compõem a Amostra B terem sido escritos por mais do que um autor;
6. Porém, considerando a consistência e a semelhança dos marcadores linguísticos nas duas amostras, a probabilidade de autoria coletiva dos textos de cada uma das amostras é idêntica, isto é, os textos da Amostra B possuem um grau de probabilidade de autoria coletiva idêntico aos textos da Amostra A.

2. Investigação linguística

2.1. Documentos

Esta perícia linguística forense independente consiste em analisar dois conjuntos de textos, identificados como Amostra A e Amostra B. A Amostra A é constituída por 40 mensagens publicadas no website [xyz.net] e pelas regras de funcionamento do website, sob o nome de utilizador “[X]”. A Amostra B é composta por 42 mensagens publicadas na rede social Facebook, utilizando o nome de utilizador “[X 1]”.

A informação relativamente à origem de cada um dos textos foi fornecida pelo Gabinete Cibercrime:

Textos que compõem a Amostra A

Texto descritivo das regras de funcionamento do website [xyz.net] Trata-se de um texto composto por cerca de 20 frases (separadas sobretudo por reticências) com três regras e conselhos destinados aos utilizadores do website. Apesar de delinear

regras, este breve texto apresenta uma escrita informal, característica dos meios de comunicação eletrónica.

Mensagens de texto publicadas no website “[xyz.net]” com comentários feitos pelo gestor em resposta a questões de vários utilizadores. São, na sua maioria, mensagens curtas destinadas aos utilizadores do website, apresentando uma escrita informal, característica dos meios de comunicação eletrónica.

Textos que compõem a Amostra B

Conjunto de 42 mensagens de texto publicadas numa conta na rede social Facebook, com o utilizador [X¹]. Estas mensagens incluem o anúncio de disponibilização de novos conteúdos no website [xyz.net], bem como mensagens de resposta a comentários de outros utilizadores da rede social em interação com o utilizador [X¹]. Estas mensagens apresentam uma escrita informal, como é comum no meio eletrónico da rede social Facebook, e, por conseguinte, menos cuidada, a exemplo do que acontece com os textos que integram a Amostra A.

2.2. Análise linguística forense de autoria

A análise realizada nesta perícia linguística assenta numa leitura exaustiva dos textos, no sentido de procurar identificar características ou elementos idiossincráticos (isto é, marcadores linguísticos) no estilo de escrita utilizado no texto descritivo das regras de funcionamento do website “[xyz.net]” e nas mensagens assinadas pelo utilizador “[xyz.net]”, publicadas no mesmo website, e que possam indiciar a autoria dos textos suspeitos. O objetivo da análise é confirmar ou infirmar a suposição de que os textos dos dois conjuntos foram escritos pelo mesmo autor. Uma análise deste tipo começa por uma leitura dos documentos que constituem a Amostra A, no sentido de identificar nos textos cuja autoria é conhecida marcadores de estilo de escrita suficientemente consistentes para demonstrarem que não se trata de uma variação ocasional ou acidental face à norma, e suficientemente distintivos para permitirem distinguir textos de autores diferentes. O mesmo método de análise é aplicado aos textos de autoria suspeita (Amostra B), sendo, subsequentemente, realizada uma comparação dos marcadores mais distintivos, no sentido de confirmar ou infirmar a suspeição de mesma autoria dos dois conjuntos de textos.

Ao considerar os critérios de consistência, é importante ter em consideração as diferenças inerentes ao género de texto existentes entre os documentos que compõem o conjunto de textos da Amostra A, bem como entre os textos da Amostra A e os textos de autoria suspeita (textos que compõem a Amostra B). Este fator constitui, frequentemente, uma situação problemática. Se os textos não partilharem características de género textual de intenção comunicativa, modo de produção e público idênticos (fatores determinantes no estilo de escrita adotado), então a análise das semelhanças e diferenças do estilo de autoria torna-se mais difícil. É, também, importante considerar aspetos quantitativos, nomeadamente volume de texto disponível para análise, uma vez que volumes de texto muito reduzidos dificultam e podem condicionar o processo de análise. Na presente análise, estes problemas não se colocam. Por um lado, os textos que integram as duas amostras pertencem a um mesmo género textual, partilhando intenção

comunicativa, modo de produção e público idênticos, e adicionalmente não apresentam uma variação significativa relativamente à sua função, nível de formalidade e proximidade face ao conteúdo. Por outro lado, o volume de texto das duas amostras é idêntico e suficiente para efetuar uma análise sustentada.


A investigação científica aplicada a este tipo de análise estilística de autoria propõe um enfoque em elementos linguísticos de base (Eagleson, 1994; McMenamin, 1993, 2002) e na relação entre pontuação e estruturas sintáticas (Chaski, 2001, 2004), bem como inicialismos e ortografia, caraterísticos dos meios de comunicação eletrónica (Sousa-Silva et al, 2011). Estes estudos mostram que os elementos linguísticos deste tipo são aqueles que são mais robustos e, por conseguinte, indicativos do estilo de autoria.

2.3. Análise

Considerou-se, após análise, que existem 11 conjuntos de características linguísticas mais distintivas nos textos publicados pelo utilizador “[X]” (Amostra A). São eles:

1. Utilização de maiúsculas e minúsculas.
2. Utilização de acentuação.
3. Gramática.
4. Utilização de abreviaturas.
5. Pontuação e espaçamento junto à pontuação.
6. Utilização de *smileys*.
7. Utilização de diminutivos.
8. Utilização de expressões do português do Brasil.
9. Utilização de numeração.
10. Erros ortográficos distintivos.
11. Escolhas lexicais idiossincráticas.

A distribuição destas características pelos textos é apresentada na Tabela 1. **S** (sim) indica a presença da característica linguística; **N** (não) indica a ausência da característica linguística. Para uma melhor identificação dos marcadores na análise realizada, as cores e os símbolos apresentados na descrição de cada marcador correspondem àqueles que são utilizados na análise.

| Marcador linguístico | Amostra A | Amostra B |
|--|-----------|-----------|
| 1. Utilização de maiúsculas e minúsculas: | | |
| Início de frase com letra minúscula | S | S |
| Início de frase com letra maiúscula | S | S |
| Utilização de letra minúscula em nomes próprios (e.g. “Portugal”, “Facebook”). | S | S |
| 2. Utilização de acentuação: | | |
| Ausência de acentos onde deveria ser utilizada acentuação (incluindo nas mesmas palavras) | S | S |
| Utilização de acentuação correta | S | S |
| Utilização de acentuação na terceira pessoa do indicativo do verbo “ser” (é, são, será, serão) | S | S |
| 3. Gramática: | | |
| Falta de preposições (“filmes comedia” em vez de “filmes de comedia”).  | S | N |

| | | |
|---|---|---|
| Contração da preposição “para” > pra | S | N |
| Confusão entre “a” de contração “a + a = à”) e a forma do verbo haver (“há”) | S | S |
| Confusão entre verbo com pronome impessoal e conjuntivo (“tornasse” > “torna-se”, “chega-se” > “chegasse”) | S | N |
| Erro de concordância em género e número: “o sua”, “opções são os melhores” “a despesas”, “uma curiosidades”, “qualquer comentarios” | S | N |
| Confusão entre “porque” > “por que” | S | N |
| Utilização da expressão “Ter [e suas conjugações] que” em vez de “Ter [e suas conjugações] de” | S | S |
| 4. Utilização de abreviaturas | S | S |
| 5. Pontuação e espaçamento junto à pontuação: | | |
| Utilização de reticências com três pontos | S | S |
| Utilização de reticências com quatro pontos | S | S |
| Utilização de reticências com cinco pontos | S | S |
| Utilização de reticências com espaço após texto precedente mas sem espaço antes do texto seguinte (e.g. [espaço][reticências com três ou quatro pontos][texto]) | S | S |
| Utilização de reticências sem espaços entre duas palavras | S | S |
| Utilização de vírgula sem espaço entre duas palavras | S | S |
| Existência de reticências em final de frase | S | S |
| Ausência de pontuação no final de frase | S | S |
| Existência de pontuação em final de frase [espaço][ponto final][texto] | S | N |
| [espaço][vírgula][texto] | S | S |
| Vírgula junto à palavra, seguida de espaço | N | S |
| Ausência de espaços junto ao sinal de parênteses curvos/retos | S | S |
| Existência de espaços junto ao sinal de parênteses (curvos/retos) | N | S |
| Dois pontos horizontais em final de frase | N | S |
| Ausência de espaços junto a pontuação (e.g. “Atão? também”, pontos finais, “[“, “:.”) | S | S |
| 6. Utilização de smileys: | | |
| Utilização de smileys simples - :) | S | N |
| 7. Utilização de diminutivos | S | N |
| 8. Utilização de expressões de português do Brasil | S | S |
| 9. Utilização de numeração: | | |
| Utilização de números árabes | S | S |
| Utilização de numerais | S | N |
| 10. Erros ortográficos distintivos: | | |
| “retficada” > “retificado”; “retficar” > “retificar”; “vistantes” > “visitantes” | S | S |
| “bloqueos” > “bloqueios” | S | N |
| “teram” > “terão”; “seram” > “serão” | S | S |
| “hora” > “ora”; “a > há” | N | S |
| 11. Escolhas lexicais idiossincráticas | | |
| Utilização idiossincrática da palavra “reeniciar” (e derivadas) | S | S |
| Utilização da expressão “mais cedo ou mais tarde” | S | N |
| Utilização das expressões “down” e “on” em vez de “down” e “up” | S | N |

Tabela 1: Distribuição do conjunto de marcadores linguísticos nas Amostras A e B.

Nem todos os marcadores linguísticos pertencentes a estes 11 grupos estão presentes em grande quantidade nos textos que compõem a Amostra A e a Amostra B; porém, quando utilizados em conjunto com outros elementos linguísticos, descrevem um estilo altamente distintivo. A existência de um conjunto reduzido destas características pode ser considerada, ainda, um estilo distintivo.

Considerações relativas aos textos

O conjunto de textos que compõem a Amostra B inclui quatro textos (sinopses de filmes e notícias) que foram reproduzidos literalmente de outros sites, nomeadamente: (“The Wicked”, página 23), reproduzido do YouTube; (“Comissão Europeia encomendou estudo sobre a partilha”, página 25), reproduzido de TekSapo; (“A Origem dos Guardiões (2012)”, página 41), reproduzido de <http://www.culturavibra.com/cinema?id=596>; (sinopse sobre Anna Karenina, página 57), reproduzido de <http://www.adorocinema.com/filmes/filme-191856>. Uma vez que estes textos são reproduzidos de outros websites, não foram incluídos na análise dos textos assinados como “[X¹]”.

A utilização errática de alguns elementos linguísticos pertencentes a um mesmo conjunto de marcadores, que representa uma aparente inconsistência entre algumas das formas linguísticas utilizadas dentro desse mesmo conjunto, pode ser indicativa de autoria coletiva, ou então pode ser justificada pelo facto de os textos se enquadrarem num género textual relativamente informal, e que permite que o(s) autor(es) alterne(m) entre formas linguísticas diferentes. Assim, a análise realizada teve em conta, não só a utilização dos marcadores linguísticos elencados, mas também a sua combinação dentro da mesma amostra. Esta análise demonstrou que existem vários marcadores linguísticos e conjuntos de marcadores nos documentos suspeitos partilhadas por textos das duas Amostras, A e B, demonstrando que aquela utilização errática é consistente entre as duas amostras de textos.

3. Opinião pessoal

Opinião relativa à autoria do conjunto de textos suspeitos

Considerando a comparação entre os textos que compõem os dois conjuntos de textos (Amostra A e Amostra B):

1. O conjunto de textos que compõem a Amostra A é estilisticamente consistente com o conjunto de textos suspeitos que compõem a Amostra B;
2. Os dois conjuntos de textos partilham 26 marcadores linguísticos distintivos, ou seja, partilham as mesmas características linguísticas altamente distintivas;
3. Com base na prova linguística, e tendo em consideração o conjunto de textos fornecido para análise, o autor dos textos da Amostra A poderá ser considerado o potencial autor dos textos que integram a Amostra B;
4. A prova linguística, isoladamente, não permite identificar o autor dos textos da Amostra A como único possível autor dos textos que compõem a Amostra B; porém, a prova linguística permitiria identificar o autor mais provável entre um conjunto de possíveis

autores.

5. Não existe qualquer prova estilística que permita excluir a possibilidade de os textos que compõem a Amostra B terem sido escritos por mais do que um autor;
6. Porém, considerando a consistência e a semelhança dos marcadores linguísticos nas duas amostras, a probabilidade de autoria coletiva dos textos de cada uma das amostras é idêntica, isto é, os textos da Amostra B possuem um grau de probabilidade de autoria coletiva idêntico aos textos da Amostra A.

Embora esta perícia assente na minha opinião cuidada, baseada na análise dos textos que me foram fornecidos, os elementos linguísticos de prova desta análise não possuem o valor inequívoco inerente a uma análise de ADN. Por conseguinte, deverão ser sempre considerados complementarmente a outras provas existentes, de natureza forense, aplicáveis ao caso em apreço, e não como possuindo em si mesma, isoladamente, valor de prova exclusiva.

4. Conclusões

Os resultados da análise realizada revelam que os textos que compõem cada uma das amostras são consistentes entre si, permitindo identificar em cada uma das amostras diversos marcadores linguísticos indicativos de autoria. Esses resultados revelam, ainda, que os marcadores mais distintivos são utilizados de forma e em quantidades semelhantes nas duas amostras de textos. A utilização alternada de diferentes marcadores linguísticos, que é, normalmente, indicativa de autoria distinta, também é idêntica nos dois conjuntos de textos, o que indicia uma escrita errática do mesmo autor ou, em alternativa, a intervenção de mais do que um autor. Esta segunda hipótese, ainda que reforçada pelo facto de algumas mensagens, em ambas as amostras, se encontrarem escritas em Português do Brasil (apesar de as restantes mensagens estarem maioritariamente escritas em Português Europeu), não é incompatível com o facto de se tratar de um mesmo autor, uma vez que, mesmo as mensagens escritas em Português do Brasil incluem elementos linguísticos característicos do Português Europeu. Do ponto de vista linguístico, poderá, assim, afirmar-se que existe uma probabilidade elevada de a autoria do conjunto de textos das duas amostras, A e B, ser a mesma.

O Perito:
Rui Sousa-Silva

Data: 30/07/2015

Este trabalho foi parcialmente apoiado por Bolsa SFRH/BPD/100425/2014 FCT-Fundação para a Ciência e Tecnologia, Portugal, co-financiada pelo POCH/FSE.



Apêndice 1 **Referências**

- Chaski, C. (2001). Empirical evaluations of language-based author identification techniques. *Journal of Forensic Linguistics: The International Journal of Speech Language and the Law*, 8(1).
- Chaski, C. (2004). Recent validation results for the syntactic analysis method for author identification. Paper presented at the Conference on Forensic Linguistics and Law, Gregynog Hall, Tregynon, Wales.
- Eagleson, R. (1994). Forensic analysis of personal written texts: a case study. In J. Gibbons (Ed.), *Language and the Law*. London: Longman.
- McMenamin, G., R. (1993). *Forensic Stylistics*. London: Elsevier.
- McMenamin, G., R. (2002). *Forensic Linguistics - Advances in Forensic Stylistics*. Boca Raton, Florida, USA: CRC Press.
- Sousa-Silva, R., Laboreiro, G., Sarmiento, L., Grant, T., Oliveira, E. & Maia, B. (2011). 'twazn me!!! ;(' Automatic Authorship Analysis of Micro-Blogging Messages. In Muñoz, R., Montoyo, A. & Métails, E. *Lecture Notes in Computer Science*, 6716.