

O estilo do crime: análise de texto em estilística forense

Rita Marquilhas^{*} & Adriana Cardoso^{**}

^{*}Faculdade de Letras da Universidade de Lisboa/CLUL

^{**}Escola Superior de Educação de Lisboa/CLUL

Abstract

Text authorship identification is one of a number of techniques developed by forensic linguistics, a discipline that uses linguistic analysis to provide evidence that can be used in the dispensation of justice. In forensic stylistics, text analysts try to characterise the choices made in particular texts as unequivocally as possible, focusing upon idiosyncrasies. The safest method seems to be to combine statistical analyses (text stylometry) with the kind of stylistic research (graphic, syntactical, lexical and textual) usually used in text analysis. This paper uses this methodology to analyse a text that has circulated on the internet since 2009 under the name of journalist Clara Ferreira Alves.

Keywords: forensic linguistics, forensic stylistics, authorship identification, keyness, style-markers

Palavras-chave: linguística forense, estilística forense, identificação de autoria, originalidade, marcadores de estilo

1. Introdução

[W]here as in principle any speaker can use any word at any time, in fact they tend to make typical and individuating co-selections of preferred words. This implies that it should be possible to devise a method of linguistic fingerprint, in other words that the linguistic ‘impressions’ created by a single speaker should be usable, just like a signature, to identify them. So far, however, practice is a long way behind theory and no one has even begun to speculate about how much and what kind of data would be needed to uniquely characterize an idiolect, nor how the data, once collected, would be analysed and stored.

(Coulthard & Johnson, 2007, p. 161)

Retirámos este trecho de um manual recente de linguística forense, escrito por Malcolm Coulthard e Alison Johnson, e decidimos usá-lo para inaugurar o nosso artigo por duas razões:

- Em primeiro lugar, qualquer exercício praticado numa área de conhecimentos recente como é a linguística forense deveria traçar brevemente a história da disciplina. Como isso esgotaria um espaço de que vamos precisar para outros efeitos, optámos por deixar aqui o testemunho de uma autoridade como é o britânico Malcolm Coulthard, diretor do Centre for Forensic Linguistics da Universidade de Aston em Birmingham. As suas investigações, juntamente com as do americano Gerald R. McMenamin, publicadas em sucessivos manuais, são as que maior ajuda prestam a quem queira fazer peritagens sobre enunciados escritos no âmbito da investigação de um crime. É esse o tema do nosso artigo, e temos que reconhecer logo de início que estamos num terreno em que há pouca fundamentação empírica para as teorias que vão surgindo, como a da existência de idioletos que por si sós identifiquem a forma original como os falantes se comportam na elaboração dos seus enunciados escritos: aqui, como dizem Coulthard e Johnson, *practice is a long way behind theory* (cf. citação em epígrafe).

- Em segundo lugar, mas em consequência da observação anterior, acontece que a falta de apoio empírico na fundamentação de teorias sobre a singularidade enunciativa dos falantes obriga a que sejam sempre muito provisórias as eventuais descobertas que o investigador pense ter feito nesta área. É preferível nem sequer lhes chamar ‘descobertas’, mas sim contributos para um trabalho que está em curso e que é o da especulação sobre a quantidade, o tipo de dados, o tipo de armazenamento e o tipo de análise que são necessários para identificar idioletos: *no one has even begun to speculate about how much and what kind of data would be needed to uniquely characterize an idiolect, nor how the data, once collected, would be analyzed and stored* (cf. citação em epígrafe).

No contexto de todas estas cautelas, entendemos o presente trabalho como uma reflexão sobre a exequibilidade de peritagens sobre enunciados portugueses no sentido de lhes identificar o autor, desconhecendo nós quem os compôs, mas havendo suspeita de ter sido um indivíduo em particular. Entendemos também que não é um exercício ocioso porque se trata de demonstrar uma faceta da utilidade social da investigação em linguística e de desenvolver instrumentos para enfrentar uma realidade nova, criada pela generalização dos meios globais de comunicação por escrito.

De facto, com a divulgação de ficheiros pela internet (correio eletrónico, *twitter*, *blogs*, *facebook*), multiplicam-se os problemas criados pela fraude textual, a qual tem sobretudo três facetas: a dos textos apócrifos, a dos textos plagiados e a dos textos anónimos. Tornou-se extremamente fácil escrever fraudulentamente em nome de outros, usar fraudulentamente os textos de outros ou ocultar com má-fé o próprio nome aquando da partilha de um enunciado escrito.

A linguística já desenvolveu uma série de disciplinas que podem apoiar a investigação destas fraudes. Trata-se agora de articular os axiomas de cada uma delas e de problematizar a forma como eles se complementam no contexto deste desafio. As disciplinas em causa são sobretudo a análise do discurso, a pragmática, a sintaxe, a crítica textual, a linguística histórica, a sociolinguística e a linguística de corpus.

2. Princípios metodológicos

A investigação desenvolvida recentemente em linguística forense orienta-se por dois princípios fundamentais. Em primeiro lugar, assume-se que é possível que os falantes repitam um mesmo *estilo* em sucessivos enunciados que escrevam. Estilo, variedade individual e escolhas são termos que, em linguística forense, correspondem às seguintes definições (McMenamin, 2002, cap. 6):

Estilo: “o reflexo de uma variedade individual, ou de grupo, nos enunciados escritos”¹;

Variedade individual: “resultado das escolhas de um autor, o qual seleciona uma forma de entre um conjunto de formas possíveis”;

Escolhas: “variedade dentro de uma norma (maneiras ‘corretas’ de dizer a mesma coisa), desvios em relação a uma norma (erros) e idiossincrasias (formas específicas do autor)”.

Em linguística forense, o analista de textos deve tentar caracterizar, por conseguinte, da forma mais inequívoca possível, as escolhas presentes nos textos que examina, sobretudo as idiossincráticas. Deve, no entanto, uma vez que os seus juízos podem implicar consequências persecutórias para terceiros, ter o cuidado de observar um segundo princípio, já ético. Tem de garantir que trabalhou com o máximo grau de precaução. O especialista Ernst Frederick Kotzé sugeriu recentemente uma forma de garantir essa precaução:

Instead of attempting to find common textual characteristics, linguistic or otherwise, the analyst should start off by attempting to prove that the texts to be compared are so different that they have probably, and even certainly, been written by different authors. If, then, by stringently applying an appropriate statistical test of significance, the analyst is not able to prove that the pair of texts show up significantly different characteristics, the way is paved to put the shoe on the other foot and analyse the texts for salient correspondences.

(Kotzé, 2010, p. 187)

¹ Para uma apresentação detalhada de como a clássica noção de estilo foi recuperada pelos estudos linguísticos, cf. Coutinho (2002).

A metodologia mais segura parece ser, por conseguinte, a de se combinarem, em modalidade de triangulação, as estatísticas lexicais, as descrições qualitativas de aspetos linguísticos e as dos aspetos extralinguísticos dos documentos em análise.

3. Estudo de caso

No presente artigo, vamos analisar um caso verídico para demonstrar como se pode praticar a triangulação acima mencionada. Vamos demonstrar quão fácil é obter resultados quantitativos e qualitativos que inicialmente aparentem solidez; e quão fácil é depois abalar tal solidez quando se parte para a multiplicação de testes diversos e para o cruzamento entre eles. O difícil é conseguir que a triangulação possa finalmente indicar, com alguma margem de certeza, qual a opinião a emitir sobre a autoria de um enunciado concreto.

Não podemos usar casos que tenham originado efetiva investigação criminal porque os que conhecemos estão em segredo de justiça, mas podemos usar outros que tenham sido noticiados na imprensa, que poderiam até ter originado queixa-crime por terem envolvido fraude textual, mas que se ficaram pela denúncia pública, prontos a serem usados em laboratório académico.

Siga-se a seguinte cronologia:

A 10 de janeiro de 2009 (ou possivelmente antes) surgiu atribuído a Clara Ferreira Alves um texto com o título ‘Este é o maior fracasso da democracia portuguesa’. (Blogue RB-Estudos, SA). A autora defendeu-se muitos meses mais tarde, em crónica publicada no *Expresso* a 5 de dezembro desse ano, alegando que se tratava de uma montagem:

Nota: Circula na Net mais um FALSO texto com a minha assinatura; truncado, com frases de uma crónica publicada no Expresso, uma crítica ao sistema de Justiça. A coberto de um FALSO endereço de Gmail com o meu nome, as pessoas pensam que sou eu que envio o texto. E o texto FALSO é citado, linkado, recitado, relinkado. O FALSO texto usa termos truculentos e idiotas. Que fazer? Entregar, por ironia, o caso à Justiça? Não me parece.

Como a própria reconheceu haver ligação com um texto da sua autoria, a crónica “A justiça criminosa” que tinha publicado em outubro de 2007, não vamos aqui tratar deste caso. Apenas nos interessa que o título que a montagem ganhou ao longo da sua circulação pelos blogues é realmente uma frase dessa crónica de Clara Ferreira Alves de 2007: trata-se da frase *Este é o maior fracasso da democracia portuguesa*. O mesmo título acabou por ser adotado também por uma crónica de um outro autor, Ricardo Santos Pinto, que também publicava em janeiro de 2009.

Em 12 de janeiro de 2009, no blogue 5dias.net, Ricardo Santos Pinto, um professor de história que se assume politicamente como próximo dos partidos de esquerda, publicou “Momentos de Lucidez”, um *post* de crítica violenta a Mário Soares (cf. Anexo I).

Três meses mais tarde, em abril de 2009, um comentário de um visitante do 5dias.net avisava que o texto estava a circular sob o nome de Clara Ferreira Alves, e Ricardo Santos Pinto até respondeu prontamente ao comentário, reclamando para si a autoria do *post* (cf. Anexo II).

Em maio de 2009, Clara Ferreira Alves denunciou o caso numa nota à sua coluna do *Expresso*.

Nota: Circula na net e na blogosfera um falso texto com a minha assinatura, que aparece como tendo sido publicado no Expresso, e que é um enunciado de injúrias a Mário Soares. O falsificador, anónimo, usou o meu nome para caluniar.

Em 2011, Clara Ferreira Alves continuava a desconhecer que se tratava de um texto de autoria já clarificada. Em 16 de abril, a autora redigia uma crónica inteira dedicada aos dois casos, o dos insultos a Mário Soares e o da montagem a partir da sua crónica sobre a justiça portuguesa (cf. Anexo III).

Como avisámos acima, não há aqui caso nenhum de linguística forense porque já tudo se encontra clarificado, se bem que não suficientemente divulgado. No entanto, podemos perguntar-nos:

«Se se desconhecesse a autoria da crónica sobre Mário Soares, publicada originalmente com o título *Momentos de lucidez*, a linguística forense poder-nos-ia ajudar a provar que ela não tinha sido escrita por Clara Ferreira Alves?».

«E se se desconfiasse do colaborador do 5dias.net enquanto possível autor desse texto, havia maneira de provar que os *Momentos de lucidez* obedecem ao seu estilo individual?»

Pegámos no caso e investigámo-lo. Concluímos que era possível encontrar contraste entre o estilo de Ricardo Santos Pinto e o de Clara Ferreira Alves, mas que a ligação entre os *Momentos de lucidez* e o estilo daquele autor não era muito óbvia. Mais importante do que a evidência linguística, seria aquela informação crucial: a de que a crónica que suscitava dúvidas de autoria tinha saído no *Expresso*. Com efeito, uma busca exaustiva no arquivo do *Expresso on-line* resultaria na conclusão de que esse texto não o integra nem nunca o integrou.

3.1. Análise quantitativa

Aplicando os procedimentos mais testados em estilística forense, escolhemos, em primeiro lugar, amostras textuais que ilustrassem de forma inequívoca o estilo de Clara

Ferreira Alves e do escritor ‘suspeito’, Ricardo Santos Pinto. Para garantir uma base homogénea, escolhemos amostras com data próxima da crónica ‘Momentos de Lucidez’, publicada em 2009. A crónica em si, por suscitar dúvidas (se bem que académicas) em relação à respetiva autoria, foi denominada de documento Q.

A amostra do estilo de Clara Ferreira Alves foi constituída pelo método da amostragem sistemática: a primeira crónica de cada mês publicada no *Expresso* em 2009 (c. 10 K palavras). Por ser a base da primeira comparação a que iríamos proceder, denominámos o ficheiro TXT que continha o conjunto das 12 crónicas de documento CFA. Já a amostra do estilo de Ricardo Santos Pinto foi constituída pela seleção de todos os seus *posts* que não contivessem trechos citados e que tivessem sido publicados no blogue 5dias.net entre meados de janeiro e inícios de março de 2009 (em março desse ano saiu do blogue e só regressou em 2011). Também deu um conjunto de c. 10K palavras, e demos ao ficheiro TXT que as continha o nome de documento RSP.

A homogeneização em formato TXT é uma condição para que funcionem sem problemas os programas de estatística lexical. Usámos o WordSmith Tools desenvolvido por Mike Scott desde 1997 por ser o que mais avançado está em termos de operações de contabilização de palavras-chave (Scott, 2005).

O WordSmith Tools permite:

- a) extrair todas as palavras individuais que ocorrem num texto e listá-las, quer alfabeticamente, quer por ordem de frequência (listas de palavras);
- b) listar e marcar graficamente palavras e cadeias de palavras dentro do contexto em que ocorrem no texto (concordâncias);
- c) listar as palavras características de um texto em comparação com as de outro - *palavras-chave* - acompanhadas da indicação do grau de originalidade, ou saliência (em inglês, *keyness*), que atingem.

Em termos de interpretação dos resultados, seguimos um princípio que tem vindo a ser defendido recentemente tanto no estudo discursivo de corpora (uma articulação entre a análise do discurso e a linguística de corpus), como em psicologia social, e, consequentemente, também em estilística forense. Trata-se de considerar que as palavras de classe fechada (determinantes, pronomes, preposições, conjunções, verbos auxiliares e alguns advérbios, os primitivos) e a sua distribuição num texto fornecem uma base sólida para o estudo quer de discursos especializados, quer de estados psicológicos, quer de estilos de escrita individual.

Closed-class keywords can form a valid and even preferable basis for empirical linguistic research into specialized discourses. (Groom, 2010, p. 59)

Relying on computerized text analysis procedures, we are finding that the examination of often-overlooked "junk words" - more formally known as function words or particles - can provide powerful insight into the human psyche. (Chung & Pennebaker, 2007, p. 344)

[...] some significant differences between documents written by the same author are to be expected, particularly if lexical words are considered. Significant differences between the use of grammatical words in different documents, however, would be indicative of different grammatical vocabularies, in particular if a similar style of writing is used [...]. (Kotzé, 2010, p. 188)

Procedemos então a uma análise contrastiva das palavras-chave nos documentos CFA, Q e RSP. Tanto na observação Q vs CFA como na observação Q vs RSP, o que nos interessou sempre foi testar duas coisas:

- a ocorrência de palavras distintivas dentro dos documentos;
- a possibilidade de os pares de documentos em análise terem um autor comum.

Usou-se o programa WordSmith Tools para se obterem, em primeiro lugar, as listas de todas as palavras que ocorriam nos três documentos em análise: as listas receberam os nomes CFA.LST, Q.LST e RSP.LST. Dentro dos grupos CFA e RSP, também se obtiveram subconjuntos de listas de palavras, destinados a testar o grau de variação interna do estilo de cada um deles.

De seguida, por meio da aplicação automática de um teste de probabilidade, determinou-se a originalidade (ou *keyness*) de cada palavra que tivesse sido identificada como ocorrendo nas listas em causa. Neste caso, “originalidade” significa “diferença significativa”. É uma análise que toma em consideração a frequência de cada palavra (o número de vezes que ela apareceu escrita) num documento e a percentagem do texto, como um conjunto, representada por aquela palavra específica. Pelo recurso a um teste de probabilidade (o do *qui quadrado*) obtêm-se as diferenças significativas entre documentos cujas frequências absolutas e percentagens de palavras tenham sido entretanto apuradas.

Considere-se, a título de exemplo, a comparação interna das crónicas de Clara Ferreira Alves apresentada no Quadro 1 (neste caso, a primeira e a segunda crónicas de 2009, sendo que a segunda serve de corpus de referência):

Ordem	Palavra-chave	Freq.	%	CR. Freq.	CR. %	Originalidade	Valor p
1	NÃO	30	3,246753	8	1,010101	8,846433	0,00294
2	CHEIRO	9	0,974026	0		6,000386	0,0143
3	NEM	8	0,865801	0		5,149753	0,02325
4	LUA	7	0,757576	0		4,304122	0,03802
5	ERA	7	0,757576	0		4,304122	0,03802
6	DE	22	2,380952	50	6,31331	-15,44025	8,5E-05

Quadro 1: Comparação interna das crónicas de Clara Ferreira Alves

A fórmula que se usou, conforme referido acima, foi a do *qui quadrado*, e a margem de erro que se estabeleceu foi a de 5%. O valor p indica isso mesmo, a fiabilidade da originalidade apurada. Qualquer originalidade com um valor p superior a 5% (i.e., 0,05) passa a ser uma originalidade negativa (caso da preposição *de*). Ou seja, quanto mais baixo for o valor p , maior é a evidência de que a *keyness*, ou originalidade, é significativa.

Os programas de análise lexical automática não distinguem entre palavras de classe fechada e de classe aberta. Tem de ser o linguista a proceder manualmente a essa separação, processo ao longo do qual surge um problema inultrapassável, criado pela metodologia da análise guiada por um corpus ao qual falte anotação linguística: o das contrações de preposição e determinante (*da*, *nos*, *às*, etc) e o das palavras homógrafas (*a*). No entanto, como se trata de formas gráficas presentes em todos os corpora, pode confiar-se na representatividade da sua distribuição contrastiva. Não se pode é extrair qualquer conclusão taxativa sobre o comportamento gramatical dos respetivos autores a partir das frequências absolutas de tais formas.

Feita a comparação interna das 12 crónicas de Clara Ferreira Alves, verificou-se que, como esperado, a originalidade das palavras de classe fechada era muito baixa, uma vez que se trata de uma amostra cronologicamente homogénea de 12 textos do mesmo género literário e escritos por uma mesma autora (média de originalidade de 9,3).

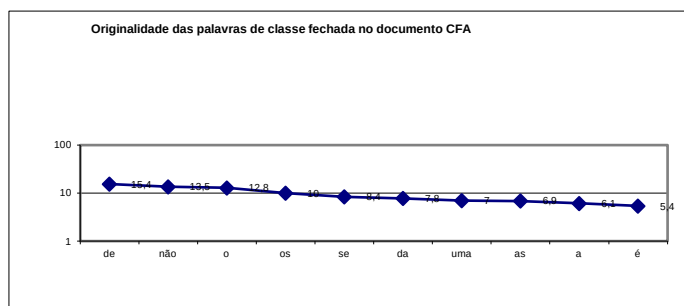


Figura 1: Riqueza do estilo de Clara Ferreira Alves

Já a originalidade das palavras de classe fechada dos mesmos textos, mas apurada agora em contraste com um corpus de referência de grande envergadura, o CRPC² (subcorpus escrito com textos da imprensa periódica, de literatura, de livro técnico-didático e de textos jurídicos – c. 8 M de palavras) atinge valores completamente diferentes (a mais baixa – *isto* - atinge uma originalidade que é mais do dobro daquela que era mais alta em termos de comparação interna – *de*; a média de originalidade é já de 104,1):

² *Corpus de Referência do Português Contemporâneo* (disponível em: www.clul.ul.pt).

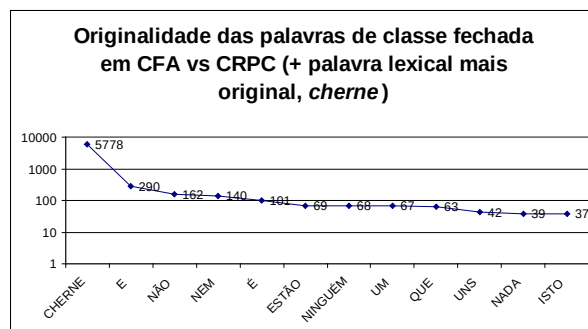


Figura 2: Singularidade do estilo de Clara Ferreira Alves

O documento RSP, contendo 10 mil palavras de 23 *posts* de Ricardo Santos Pinto, em termos de comparação interna, revela uma originalidade com valores inferiores aos que se encontram em Clara Ferreira Alves (média de originalidade de 6,1):

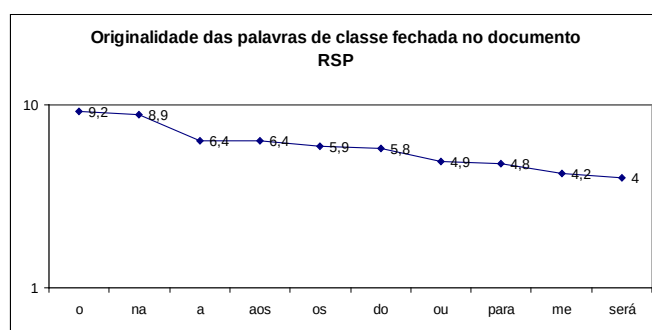


Figura 3: Riqueza do estilo de Ricardo Santos Pinto

Começa a formar-se a conclusão de que Clara Ferreira Alves parece uma autora mais singular e rica do que Ricardo Santos Pinto no que diz respeito aos vocabulários gramaticais de ambos. Essa impressão confirma-se no contraste do documento RSP com o CRPC. Os valores de originalidade das palavras gramaticais de Ricardo Santos Pinto são mais uma vez inferiores aos de Clara Ferreira Alves, em torno da média de 74,6 (contra 104,1 nela, lembre-se), demonstrando que o estilo do autor se destaca menos daquela mediania escrita que o corpus de referência de 8 milhões de palavras pretende representar:

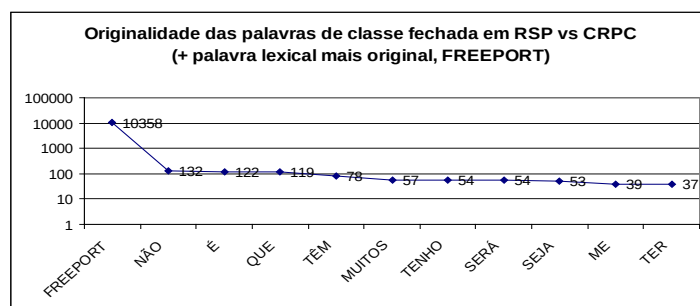


Figura 4: Singularidade do estilo de Ricardo Santos Pinto

Estes dois gráficos de contraste com o CRPC demonstram adicionalmente a especificidade do género literário em causa, ou seja, a crónica jornalística de cariz político. O estilo é coloquial e a retórica é argumentativa. Nestas palavras-chave, há uma proximidade aos corpora da interação oral, que também têm *é* e *não* entre as palavras de classe fechada com uma originalidade mais alta (teste feito contrastando o CORDIAL-SIN³ com o mesmo subcorpus escrito do CRPC). Por outro lado, dedicando-se sobretudo os autores de crónicas a denunciar estados de coisas e a usar uma lógica argumentativa recheada de premissas, constantemente referem os atributos presentes (*é*) e ausentes (*não* X) em entidades nomeadas no universo de referência. A título ilustrativo, vejam-se as colocações de *é* nos documentos CFA e RSP:

N	Left5	Left4	Left3	Left2	Left1	Centre	Right1	Right2	Right3	Right4	Right5
1	QUE	QUE	E	O	QUE	É	UM	É	E	O	E
2	E	O	A	DE	NÃO		O	NÃO		QUE	
3			DO	É			QUE			A	
4							UMA			E	
5							A				

Quadro 2: Colocações de *é* em CFA

N	Left5	Left4	Left3	Left2	Left1	Centre	Right1	Right2	Right3	Right4	Right5
1	QUE	O	QUE	SERÁ	QUE	É	QUE	QUE	PARA	QUE	NÃO
2		NÃO	DE	O	NÃO		UM		A	O	
3							MELHOR		DE	DE	
4							O			A	
5							VERDADE				

Quadro 3: Colocações de *é* em RSP

Temos portanto nesta análise, até agora indutiva porque guiada pelo conteúdo dos corpora e pela estatística sobre as suas frequências, uma primeira base empírica para afirmar que os recursos estilísticos de Ricardo Santos Pinto se distinguem dos de Clara Ferreira Alves por serem menos originais, se bem que ambos os autores pareçam seguir um mesmo padrão textual.

A comparação que falta fazer é entre o documento Q e as amostras do estilo dos dois autores. Qual é o vocabulário gramatical significativo, original, de Q, comparado com os documentos CFA e RSP?

³ *Corpus Dialectal para o Estudo da Sintaxe / Syntax-oriented Corpus of Portuguese Dialects* (disponível em: www.clul.ul.pt).

Com Referência a CFA		Com Referência a RSP	
Palavras	Originalidade	Palavras	Originalidade
LHE	191	LHE	264
A	13	A	19
QUE	10	QUE	6
HOJE	5	NUM	4
NESSAS	4	UM	2,7
APÓS	4	---	---
NUM	3	---	---

Quadro 4: Originalidade de palavras de classe fechada no documento Q

Poucas conclusões se poderiam tirar deste quadro se não se tivesse procedido ao estudo anterior, que indicou o estilo de Ricardo Santos Pinto como menos original e variável que o de Clara Ferreira Alves. Aparentemente, Q até se aproxima mais, nas suas palavras-chave de classe fechada dos textos da autora. Só que é preciso ver, em primeiro lugar, que a expressão que anaforicamente (em sentido retórico) inicia todos os sucessivos sintagmas nominais, i.e., ‘a lucidez que *lhe*...’, estratégia ausente de todos os outros textos aqui em análise, torna aconselhável que se excluam da avaliação as palavras gramaticais *a*, *que* e *lhe*, precisamente as que têm mais originalidade nos dois testes. A partir daí, sobram só seis palavras, as quais apresentam um baixo grau de originalidade contra ambos os corpora de referência. De qualquer forma, a distanciação em relação às crônicas de Clara Ferreira Alves é maior do que em relação aos *posts* de Ricardo Santos Pinto (quatro palavras chave contra duas, e médias de 4 e 3,4, respetivamente). Além disso, como se tinha já visto que aquele autor variava pouco o seu estilo gramatical, o facto de a originalidade de Q ser muito baixa torna mais razoável que, havendo que escolher entre os dois, se escolha antes Ricardo Santos Pinto como o responsável pela autoria do documento em análise.

3.2. Análise qualitativa

A análise qualitativa de um estilo centra-se nas formas e estruturas linguísticas usadas por um autor:

The focus of the qualitative study of writing is a systematic linguistic description of what forms are used by a writer, and how and why they may be used (McMenamin, 2002, p. 173)

A metodologia usada na análise qualitativa envolve as seguintes etapas: (i) observação sistemática de corpora; (ii) identificação de marcadores de estilo; (iii) descrição; (iv) comparação; (v) conclusões.

Os diferentes níveis de análise linguística (sintaxe, léxico, estrutura textual, pontuação, ortografia) podem funcionar como marcadores de estilo. A relevância de cada um dos níveis vai contudo depender de vários fatores, tais como o suporte de escrita e o uso de *software* específico. Assim, caso o texto seja redigido em computador, a ortografia passa a ser um nível de observação pouco fiável para a determinação de autoria. Tal acontece porque, se o texto estiver ortograficamente correto, essa propriedade pode ter resultado, por exemplo, da ativação de um corretor automático.

Mas, ao contrário da ortografia, e devido à pouca eficácia dos corretores automáticos do português, a pontuação não é sistematicamente corrigida pelo programa de computador, a não ser em casos muito específicos, como o do ponto final seguido de letra maiúscula. Trata-se desta forma de um domínio em que mais facilmente se encontram indícios de semelhança autoral.

Por outro lado, as estruturas sintáticas constituem marcadores bastante importantes para a determinação de autoria, dado que o ‘uso fraudulento’ da sintaxe exige um conhecimento explícito do funcionamento da língua que não é facilmente acessível a um autor não familiarizado com os estudos linguísticos. Em termos práticos, esta observação significa que um autor leigo em linguística, ao redigir um documento ‘em nome de outrem’, dificilmente se conseguirá distanciar de forma credível e sistemática das estruturas sintáticas presentes nas suas produções escritas (cf. Chaski, 2001).

Urge ainda referir que, na análise qualitativa desenvolvida pela linguística forense, a singularidade enunciativa não é estabelecida habitualmente pela presença de um marcador de estilo, mas sim pela coexistência de vários marcadores nos mesmos grupos de texto.

No texto Q em análise, foram identificados os seguintes marcadores de estilo:

Pontuação

- Uso de aspas «»
- Uso de reticências

Sintaxe/pontuação

- Ocorrência de grupos nominais marcados graficamente como frases

3.2.1. Uso de aspas

O contraste existente nos textos em análise relativamente ao uso de aspas é sistematizado no Quadro 5:

Q	CFA	RSP
« »	“ ”	« », “ ”

Quadro 5: Uso de aspas nos diferentes grupos de texto

Em Q, só ocorrem aspas (e nunca aspas altas):

- (1) A lucidez que lhe permitiu conduzir da forma «brilhante» que se viu o processo de descolonização.
- (2) A lucidez que lhe permitiu governar sem ler os «dossiers».
- (3) A lucidez que lhe permitiu, após a vitória nessas eleições, fundar um grupo empresarial, a Emaudio, com «testas de ferro» no comando
- (4) A lucidez que lhe permitiu silenciar, através de pressões sobre o director do «Público», José Manuel Fernandes
- (5) A lucidez que lhe permitiu considerar José Sócrates «o pior do guterrismo»
- (6) A lucidez que lhe permitiu ler os artigos «O Polvo» de Joaquim Vieira na «Grande Reportagem»,

Nos textos de Clara Ferreira Alves, só ocorrem aspas altas (e nunca aspas):

- (7) interpretando a vitória do PS como mais um fenómeno da "sorte" de Sócrates (CFA)
- (8) Seria bom que a sibila se deixasse de pronunciamentos vagos, apagasse a fogueira da instabilidade e se comportasse como o primeiro magistrado da nação, refém da unidade nacional e não de intrigas e "manobras pífias", para citar o desassossegado director do "Público" (CFA)
- (9) O filme de Michael Moore, "Capitalismo, Uma História de Amor", é muito claro (CFA)

Nos textos de Ricardo Santos Pinto, o uso de aspas (« ») e aspas altas (“ ”) encontra-se em distribuição complementar: as aspas são usadas para destacar palavras, para indicar discurso direto, título blogues e de *posts*, título de livros, nomes de restaurante e para citações; as aspas altas, por sua vez, são usadas nos mesmos contextos, mas no interior de uma citação:

- (10) Ainda sobre o mesmo assunto, e após um texto de João Pinto e Castro a defender, como sempre, a Ministra, Rui Tavares refere que «insistir numa reforma apenas porque é "impopular" é uma desculpa fácil. Difícil é fazer uma reforma compreensível e motivadora para quem vai ter de participar nela. Mas às vezes é possível, e nesses casos é essencial.» (RSP)

- (11) Paulo Pinto aborda o estado do Sistema Nacional de Saúde e as burocracias que o rodeiam: « Por isso, da próxima vez que me vierem falar em simpleses, em confiança entre os cidadãos e o Estado, em reformas simplificadoras, em racionalização, no raio que os parta, não aceitarei tangas de orçamentos, taxas, estatísticas, milhões e outros grandes paleios da treta. Se alguém me pedisse opinião, sugeriria apenas “correção” e “bom-senso”.» (RSP)

Com base no uso de aspas, conclui-se assim que parece haver maior semelhança autoral entre Q e o grupo de textos RSP do que entre Q e o grupo de textos CFA. O facto de Q não exibir aspas altas pode dever-se simplesmente à inexistência de contextos que envolvam o uso de aspas no interior de citação.

Note-se, porém, que a análise deste tipo de marcadores de estilo exige alguma cautela por parte do investigador, dado que o uso de aspas/aspas altas pode ser determinado pela política editorial das diferentes publicações. Assim, poder-se-ia pensar que Clara Ferreira Alves só usa aspas altas por imposição do livro de estilo do *Expresso*.

Neste caso, porém, tal corroboraria a análise aqui proposta. Q “aparece como tendo sido publicado no *Expresso*” (CFA). Ora tal proposta seria infirmada precisamente pelo facto de este texto não seguir as normas de estilo da referida publicação.

3.2.2. Uso de reticências

A quantificação do uso de reticências nos três grupos de texto é apresentada no Quadro 6:

Q	CFA	RSP
3	7	7

Quadro 6: Uso de reticências nos diferentes grupos de texto

Tendo em conta a posição das reticências na frase, estas podem ocorrer em posição medial ou em posição final de frase; veja-se a quantificação apresentada no Quadro 7:

	Q	CFA	RSP
posição final	0	4	5
posição medial	3	3	2

Quadro 7: Uso de reticências nos diferentes grupos de texto

Em termos quantitativos, a distribuição das reticências em CFA e RSP é semelhante. Contudo, o valor estilístico associado ao uso de reticências não é equivalente.

Considerando apenas os contextos que envolvem uma posição medial (que é a que ocorre em Q), observa-se que em CFA as reticências são utilizadas, no corpo do texto, para substituir letras ou palavras e, no título de uma crônica, para traduzir graficamente uma enumeração elidida, que logo a seguir aparece explícita na expressão ‘e do resto’.

- (12) e quando for lá de certeza que lhe ligo, ou o contrário, se vier por cá, ligue-me e vamos... preencher o pontilhado com o verbo que lhe calhar melhor. (CFA)
- (13) As respostas variaram entre isto: "Aquilo dos mariconços casados? Que pan...ice. (CFA)
- (14) A desordem dos advogados... e do resto (CFA)

Pelo contrário, no grupo de textos de RSP, as reticências ocorrem apenas no corpo do texto e têm um valor estilístico claro: o de introduzir uma nota de suspense, com valor marcadamente irônico (este valor poderia ser parafraseado por: *nem mais nem menos...*).

- (15) e sabemos também que o ministro Augusto Santos Silva acusa os professores de não saberem distinguir entre Salazar e a democracia e compara o paladino da liberdade, Mário Soares, com... Mário Nogueira (RSP)
- (16) Entretanto, o mês termina com a polémica dos contentores de Alcântara e com o incrível alargamento da concessão ao grupo... Mota-Engil de Jorge Coelho (RSP)

Ora é precisamente este valor que se encontra em duas das ocorrências de Q:

- (17) A lucidez que lhe permitiu que o Estado lhe arrendasse e lhe pagasse um gabinete, a que tinha direito como ex-Presidente da República, na... Fundação Mário Soares.
- (18) A lucidez que lhe permitiu fazer obras no Colégio Moderno, propriedade da família, sem licença municipal, numa altura em que o Presidente era... João Soares

Conclui-se, assim, que o valor estilístico associado ao uso de reticências, ainda que não possa determinar *per se* que se suspeite de uma coincidência autoral, pode contribuir, em conjunto com outros traços linguísticos, para apoiar a tese de coincidência autoral entre Q e RSP.

3.2.3. Ocorrência de grupos nominais marcados graficamente como frases

O documento Q tem uma estrutura paralelística, que se caracteriza pela repetição exaustiva da sequência “A lucidez que ...”, que ocorre 21 vezes ao longo do texto. O efeito da repetição é claro: pretende-se reforçar o elevado número de ações condenáveis de Mário Soares ao longo da sua vida.

Do ponto de vista sintático, esta repetição traduz-se na elevada ocorrência de grupos nominais constituídos por um núcleo nominal modificado por uma oração relativa restritiva:

- (19) A **lucidez** que lhe permitiu escapar à PIDE e passar um bom par de anos, num exílio dourado, em hotéis de luxo de Paris. A **lucidez** que lhe permitiu conduzir da forma «brilhante» que se viu o processo de descolonização. A **lucidez** que lhe permitiu conseguir que os Estados Unidos financiassem o PS durante os primeiros anos da Democracia. A **lucidez** que o fez meter o socialismo na gaveta durante a sua experiência governativa.

Articulando a sintaxe com a pontuação, é assim possível definir um marcador de estilo com bastante representatividade em Q: a ocorrência de grupos nominais marcados graficamente como frases, isto é, iniciados com letra maiúscula e terminados com ponto final.

Ora a análise do grupo de textos de CFA e RSP permite concluir que este é também um traço importante do estilo dos dois autores. Vejam-se alguns exemplos:

- (20) Lembrei-me de crepes chineses, nem sei porquê. **Crepes chineses gigantes, enrolados em cobertores e lençóis sujos.** (CFA)
- (21) O Muro matou cerca de 200 pessoas. No mundo em que vivemos é provável que o Muro se torne apenas uma efeméride. **Uma fotografia. Uma memória.** (CFA)
- (22) Tantos anos passaram, tanto dinheiro mudou de mãos, e eles ainda ali estão, os sem-abrigo. **Vestígios arqueológicos de uma civilização ferida.** (CFA)
- (23) Na área das transmissões televisivas, a Controlinveste tem lucros anuais de 150 milhões de euros. **Uma margem obscena, conseguida à custa do sufoco dos clubes e do adiantamento de verbas em troca de mais anos de contrato.** (RSP)
- (24) Em segundo lugar, porque, apesar das divergências que culminaram na sua saída, faz parte da história do «5 Dias». **Ela e todos aqueles que saíram na mesma altura** (RSP)

Conclui-se, assim, que os grupos nominais marcados graficamente como frases, embora frequentes em Q, não são relevantes para a identificação de autoria. Esta conclusão deve-se ao facto de este marcador de estilo constituir, não uma marca idioletal, mas sim socioletal (no sentido em que parece ser partilhado por um grupo socioprofissional específico).

4. Conclusão

Neste estudo, procurámos aplicar a metodologia desenvolvida pela linguística forense para determinar a autoria da crónica *Momentos de lucidez*, cuja autoria é reclamada pelo jornalista Ricardo Santos Pinto, mas que chegou a circular na internet como sendo da autoria da jornalista Clara Ferreira Alves.

A questão central que orientou este estudo foi: «Se se desconhecesse a autoria da crónica *Momentos de lucidez*, a linguística forense poder-nos-ia ajudar a provar que esta crónica não tinha sido escrita por Clara Ferreira Alves?».

Para responder a esta questão, foram empreendidas duas análises independentes: quantitativa e qualitativa. Estas análises apontam apenas para a probabilidade de Ricardo Santos Pinto ter escrito a crónica *Momentos de lucidez*. Ainda que em linguística forense se trabalhe sempre com probabilidades e não com certezas absolutas, é de registar que neste caso concreto a impossibilidade de chegar a conclusões mais firmes se deve sobretudo a dois fatores: (i) a estrutura da crónica e (ii) a semelhança de estilo de Clara Ferreira Alves e Ricardo Santos Pinto. Assim, se, por um lado, a estrutura paralelística da crónica, que envolve a repetição de grupos nominais marcados graficamente como frase, limita a possibilidade de identificar marcadores de estilo diversificados, nomeadamente ao nível da sintaxe, o facto de ambos os jornalistas terem um estilo semelhante (e não se afastarem da norma padrão) dificulta a tarefa de definição de uma singularidade enunciativa.

Referências

- Chaski, Carole (2001) Empirical evaluations of language-based author identification techniques. *Forensic Linguistics*, 8 (1), pp. 1-65.
- Chung, Cindy & James Pennebaker (2007) The psychological functions of function words. In Klaus Fiedler (ed.) *Social Communication*. New York & Hove: Psychology Press, pp. 343–359.
- Coulthard, Malcolm & Alison Johnson (2007) *An introduction to forensic linguistics: language in evidence*. London & New York: Routledge.
- Coutinho, Maria Antónia (2002) Perspectivas linguísticas sobre a noção de estilo. In *Actas do Encontro Comemorativo dos 25 Anos do CLUP*. Porto: CLUP, 2, pp. 41–54.

- Groom, Nicholas (2010) Closed-class keywords and corpus-driven discourse analysis. In Marina Bondi & Mike Scott (eds.) *Keyness in Texts*. Amsterdam & Philadelphia: John Benjamins, pp. 59–78.
- Kotzé, Ernst Frederick (2010) Author identification from opposing perspectives in forensic linguistics. *Southern African Linguistics and Applied Language Studies* 28 (2), pp. 185–197.
- McMenamin, Gerald (2002) *Forensic Linguistics: Advances in Forensic Stylistics*. Boca Raton Fla.: CRC Press.
- Scott, Mike (2005) *WordSmith Tools 5.0*. Oxford: Oxford Univ. Press.

ANEXO I

5. Momentos de lucidez

12 de Janeiro de 2009 por Ricardo Santos Pinto

Mário Soares, num dos momentos de lucidez que ainda vai tendo, veio chamar a atenção do Governo, na última semana, para a voz da rua. A lucidez, uma das suas maiores qualidades durante uma longa carreira política.

A lucidez que lhe permitiu escapar à PIDE e passar um bom par de anos, num exílio dourado, em hotéis de luxo de Paris. A lucidez que lhe permitiu conduzir da forma «brilhante» que se viu o processo de descolonização. A lucidez que lhe permitiu conseguir que os Estados Unidos financiassem o PS durante os primeiros anos da Democracia. A lucidez que o fez meter o socialismo na gaveta durante a sua experiência governativa. A lucidez que lhe permitiu governar sem ler os «dossiers». A lucidez que lhe permitiu não voltar a ser primeiro-ministro depois de tão fantástico desempenho no cargo. A lucidez que lhe permitiu pôr-se a jeito para ser agredido na Marinha Grande e, dessa forma, vitimizar-se aos olhos da opinião pública e vencer as eleições presidenciais. A lucidez que lhe permitiu, após a vitória nessas eleições, fundar um grupo empresarial, a Emaudio, com «testas de ferro» no comando e um conjunto de negócios obscuros que envolveram grandes magnatas internacionais.

A lucidez que lhe permitiu receber do Estado, ao longo dos últimos anos, donativos e subsídios superiores a um milhão de contos. A lucidez que lhe permitiu receber, entre os vários subsídios, um de quinhentos mil contos, do Governo Guterres, para a criação de um auditório, uma biblioteca e um arquivo num edifício cedido pela Câmara de Lisboa. A lucidez que lhe permitiu receber, entre 1995 e 2005, uma subvenção anual da Câmara Municipal de Lisboa, na qual o seu filho era Vereador e Presidente. A lucidez que lhe permitiu que o Estado lhe arrendasse e lhe pagasse um gabinete, a que tinha direito como ex-Presidente da República, na... Fundação Mário Soares. A lucidez que lhe permite que, ainda hoje, a Fundação Mário Soares receba quase 4 mil euros mensais da Câmara Municipal de Leiria. A lucidez que lhe permitiu fazer obras no Colégio Moderno, propriedade da família, sem licença municipal, numa altura em que o Presidente era... João Soares. A lucidez que lhe permitiu silenciar, através de pressões sobre o director do «Público», José Manuel Fernandes, a investigação

jornalística que José António Cerejo começara a publicar sobre o tema. A lucidez que lhe permitiu candidatar-se a Presidente do Parlamento Europeu e chamar dona de casa, durante a campanha, à vencedora Nicole Fontaine. A lucidez que lhe permitiu considerar José Sócrates «o pior do guterrismo» e ignorar hoje em dia tal frase como se nada fosse. A lucidez que lhe permitiu passar por cima de um amigo, Manuel Alegre, para concorrer às eleições presidenciais uma última vez. A lucidez que lhe permitiu, então, fazer mais um frete ao Partido Socialista. A lucidez que lhe permitiu ler os artigos «O Polvo» de Joaquim Vieira na «Grande Reportagem», baseados no livro de Rui Mateus, e assistir, logo a seguir, ao despedimento do jornalista e ao fim da revista. A lucidez que lhe permitiu passar incólume depois de apelar ao voto no filho, em pleno dia de eleições, nas últimas Autárquicas. No final de uma vida de lucidez, o que resta a Mário Soares? Resta um punhado de momentos em que a lucidez vem e vai. Vem e vai. Vem e vai. Vai... e não volta mais.

FONTE: Blogue 5 dias.net (<http://5dias.net/2009/01/12/momentos-de-lucidez/>)

ANEXO II

julio murraças says:

15 de Abril de 2009 at 23:51

Como é que este texto circula na net como sendo um artigo da Clara Ferreira Alves no Expresso?

Não tenho simpatia política, nem pessoal, pelo visado, mas algo não cheira bem ...

[...]

Ricardo Santos Pinto says:

16 de Abril de 2009 at 0:19

Parece-me que vou ter de processar a Clara Ferreira Alves. O texto é meu e só meu, Júlio Murraças. E assumo-o da primeira à última linha.

FONTE: Blogue 5 dias.net (<http://5dias.net/2009/01/12/momentos-de-lucidez/>)

ANEXO III

Pluma Caprichosa

Este é o maior fracasso da democracia

Clara Ferreira Alves

Circula na net como um vírus um texto com o título ESTE É O MAIOR FRACASSO DA DEMOCRACIA PORTUGUESA, falsamente assinado por mim e falsamente publicado no Expresso. O texto começou a circular há cerca de dois anos e apesar dos desmentidos que entretanto fiz, tanto no programa "Eixo do Mal" como numa nota nesta coluna, o texto continua a circular e continuo a receber mails tanto de admiradores (!) do texto como de leitores que acham surpreendente o ataque a um amigo meu. A essência do texto, escrito num estilo truculento e pedestre, é a difamação de Mário Soares. Não se trata de um pequeno insulto,

porque o texto, além de ser fraudulento e usurpar cobardemente o meu nome e a minha assinatura, é vexatório e difamatório para um ex-primeiro-ministro e um ex-Presidente da República. Mário Soares é acusado de ter vendido a pátria e de ser o culpado de todos os vícios e crimes da democracia portuguesa.

Quando o texto apareceu, tanto Mário Soares como eu optámos por desvalorizar a sua existência. Não era, por essa altura, um texto tão multiplicado e expandido como hoje. Quem o pôs a circular teve o cuidado de o enviar para personalidades socialistas e foi um dirigente socialista que alertou o Expresso para o texto. Que eu desconhecia. Uma pessoa conseguiu também por essa altura traçar o texto num blogue de direita com o título ‘Nonas’, onde o texto era publicado com o meu nome. Hoje, são às centenas os blogues que o reproduzem.

Além deste texto, circula ainda outro, cujo título não recordo, também falso, em que insulto o Parlamento e chamo aos deputados e políticos “cavalgadas”. É um texto de ódio e odioso, tão pedestre como o anterior, e quem estiver habituado a ler-me detetará que nunca podia escrevê-lo. Isto aconteceu também a Miguel Sousa Tavares, que há uns anos desmentiu que um texto posto a circular fosse seu. Miguel Sousa Tavares e eu sabemos que vivemos numa terra sem lei onde o bom nome das pessoas pode ser livremente insultado e usurpado sem que se possam tomar providências. Crimes impunes a que o Ministério Público não liga.

A Internet tornou-se, juntamente com a escória que usa as caixas de comentários de jornais online para vaziar o seu rancor e a sua prosa analfabeta, um território pornográfico onde é possível matar impunemente, injuriar impunemente, vender carne humana impunemente, roubar impunemente. O algoritmo da Google transforma uma contrafação num texto fundamental de um autor, e transforma um insulto e uma falsidade numa verdade (já me aconteceu antes, com os vômitos de Vasco Pulido Valente sobre a minha pessoa). A Google arroga-se o direito de retirar da net o que quer e o que não quer e alega que só pode retirar este tipo de textos, que marcam para sempre, porque o que está na net fica na net, por ordem do tribunal. Se temos de esperar por uma ordem do tribunal podemos esperar o resto da vida. No entanto, quando se tratou de retirar do Google Images a fotografia de um chimpanzé que aparecia quando se buscava o nome de Michelle Obama, a Google cedeu. Retirou a imagem. Os protestos e pressões obrigaram a tal.

Eis como o meu nome e a minha reputação, e a de Mário Soares, estão nas mãos de criminosos e dos Pilatos da tecnologia. O que o texto pretende é desmascarar a minha infâmia, pois seria capaz de atraí-loar um amigo insultando-o daquele modo, e desmascarar a infâmia de Mário Soares, que é, segundo a prosa viral, um ladrão sem perdão.

Não me espanta que este tipo de manobras e infâmias apareça contra a área socialista. Por razões obscuras, são os socialistas em Portugal que aparecem enlameados em campanhas, rumores e insinuações. O caso Casa Pia foi disto o claro exemplo. Nada disto é inocente e o nosso sistema político escolheu lavar daí as mãos ou contribuir para prolongar a infâmia.

Este é um tempo em que prolifera o insulto anónimo e a degradação moral é a regra. O excesso e a trivialização da informação, a sua velocidade de propagação, fazem com que só a irrelevância nos preocupe e mobilize. O país, no dia em que perdeu a sua última parcela de

soberania e se tornou um protetorado dos bancos e credores, discutia com ardor a escolha de um homem que passou de cidadão a candidato político e que perdeu a sua base de apoio entregando-a a um partido, o PSD. Fernando Nobre mandou fechar a página do Facebook, para não se sujeitar aos insultos. Na verdade, podemos discordar do que julgamos ser o oportunismo de Nobre mas isso não nos dá o direito a insultá-lo alarvemente. Uma sociedade que abdica dos mecanismos de vigilância da liberdade de expressão admite tudo e perdoa tudo e as criaturas do lamaçal prosperam. Este, sim, é um dos fracassos da democracia mediática.

FONTE: *Expresso*, Revista Única, 16 de abril de 2011.