

## **ATRIBUIÇÃO DE AUTORIA TEXTUAL COM PROPÓSITOS FORENSES: PANORAMA E PROPOSTA DE ANÁLISE<sup>1</sup>**

**Dayane Celestino de Almeida<sup>2</sup>**

dayane.almeida@usp.br

The Unabomber sent us a 35.000 word essay claiming to explain his motives and views of the ills of modern society [...]. After the manifesto appeared in *The Washington Post* and *The New York Times*, thousands of people suggested possible suspects. One stood out: David Kaczynski described his troubled brother Ted [...]. (and) provided letters and documents written by his brother. Our linguistic analysis determined that the author of those papers and the manifesto were almost certainly the same. (FBI 100. *The Unabomber*)

**RESUMO:** A pergunta “Quem escreveu este texto?”, que emerge frequentemente quando se tem dúvidas sobre a autoria de textos literários, históricos ou religiosos, tem sido também feita na esfera forense, a fim de solucionar crimes ou processos judiciais. Este trabalho tem por objetivo traçar um breve panorama dos estudos de atribuição de autoria em contextos judiciais, bem como verificar problemas que se colocam com relação aos estudos atuais, propondo uma análise alternativa e/ou complementar em alguns casos.

**Palavras-chave:** Atribuição de autoria; Semiótica Discursiva; Estilo; Linguística Forense.

### **INTRODUÇÃO**

O interesse pela autoria de textos escritos é tão antigo quanto a própria escrita. Desde que o homem passou a registrar a sua língua dessa forma, podem ser muitas as situações em que emerge a pergunta “Quem escreveu este texto?”. Ela pode surgir porque os textos são anônimos, escritos por pseudônimos ou porque há uma disputa com relação a sua autoria. Inicialmente, tal pergunta fazia parte apenas do trabalho daqueles que se debruçavam sobre textos históricos, religiosos ou literários. Mais

---

<sup>1</sup> Desenvolvido com o apoio do CNPq. Vale ressaltar que este trabalho faz parte de nossa pesquisa de Doutorado, ainda em andamento.

<sup>2</sup> Universidade de São Paulo (USP).

recentemente, no entanto, saber quem escreveu um determinado texto ou conjunto de textos pode ajudar a solucionar crimes ou dirimir contendas judiciais. Nesses contextos, entra em cena o linguista forense, cujo trabalho pode elucidar tais casos na medida em que sua análise é capaz de revelar estilos linguísticos que podem ser relacionados tanto a um indivíduo quanto a um grupo social.

A Linguística Forense, em sua definição *strictu sensu*, é a aplicação da Linguística com vistas ao fornecimento de evidências em processos judiciais ou à solução de crimes. Olsson (2008, p. 6) lembra que o adjetivo *forensic*, em inglês, está associado a uma ciência capaz de resolver crimes<sup>3</sup> (no Brasil, tal papel é assumido, via de regra, pelo que se conhece por “Criminalística”). Assim, enquanto a área englobante de Linguagem e Direito inclui pesquisas que se dedicam à linguagem dos textos legais, à linguagem da polícia e dos tribunais, à interpretação e tradução legais, a Linguística Forense propriamente dita, embora faça parte dessa grande área, tem como objetos específicos a identificação de falantes, os crimes de linguagem (e.g. injúria, difamação, assédio, suborno, ameaça, extorsão)<sup>4</sup>, os textos cujo significado é “questionável” ou ambíguo, as marcas registradas<sup>5</sup> - que podem ser “disputadas” por diferentes partes (COULTHARD, 2004a; BUTTERS, 2007) -, a adequação de advertências escritas (e.g. nas embalagens de cigarro, como mostram Coulthard e Hagemeyer [2013]), a compreensibilidade de instruções escritas (BUTTERS, 2011 e, por fim, a análise de autoria textual; ou seja, a Linguística Forense se relaciona ao fornecimento de evidências ou à resolução de crimes<sup>6</sup>.

Ainda pouco difundida no Brasil, a análise de autoria com propósitos forenses já é bastante utilizada em outros países, principalmente no Reino Unido e nos Estados Unidos. De fato, já houve até mesmo alguns casos “de peso” solucionados com a sua ajuda, como é o caso do *Unabomber*, que aparece na epígrafe. Entre 1978 e 1995, nos Estados Unidos, várias bombas foram enviadas pelo correio a diferentes pessoas. Com o passar do tempo, o FBI notou que as pessoas que recebiam as bombas eram todas funcionárias de universidades ou companhias aéreas (daí o nome *Unabomber*: Un, de “University” e A, de “Airlines”). Em 1995, alguns jornais de

---

<sup>3</sup> No site *Cambridge Dictionaries Online*, encontra-se a seguinte aceção para “*forensic*”: “related to scientific methods of solving crimes, involving examining the objects or substances that are involved in the crime”. Disponível em <<http://dictionary.cambridge.org/dictionary/british/forensic?q=forensic>>. Acesso em 20/11/2012. Tradução nossa: “relaciona-se a métodos científicos para resolver crimes, incluindo o exame de objetos ou substâncias que estão envolvidas em um crime”.

<sup>4</sup> Para uma revisão de tais crimes, ver Solan e Tiersma (2005).

<sup>5</sup> Em inglês: “*Trademark*”.

<sup>6</sup> Sobre a especificidade da Linguística Forense e a abrangência da área de Linguagem e Direito: Coulthard (2012); Kredens (2012); Tiersma (s.d.); Turell (2012). Solan e Tiersma (2005) usam, inclusive, o termo “*linguistic evidence*” no lugar de “*forensic linguistics*”. Gibbons (2011) propõe o termo “*communication evidence*”.

circulação nacional receberam um manuscrito intitulado *Industrial Society and its Future* de alguém que alegava ser o *Unabomber*. O manuscrito foi publicado pelo jornal Washington Post, e alguns meses depois o FBI foi contatado por um homem dizendo que o manuscrito soava como sendo de seu irmão, um ex-professor universitário de matemática. O caso estava sendo analisado pelo agente do FBI Jim Fitzgerald, que estava lendo atentamente o manifesto em busca de pistas que pudessem levar ao seu autor. Após a denúncia, Fitzgerald comparou o manifesto com texto que o suspeito enviara a sua família no decorrer dos anos, encontrando ali padrões recorrentes, o que serviu de base para a expedição de um mandado de busca. Como resultado, o FBI localizou e prendeu Ted Kaczynski. Para corroborar a análise de Fitzgerald, o FBI contratou Donald Foster (que havia tido sucesso em alguns casos de autoria), e a análise de Foster mostrou que os textos encontrados e o manifesto compartilhavam várias escolhas linguísticas, tanto lexicais quanto gramaticais, e haviam sido escritos pela mesma pessoa. Segundo Grant (2004), esse caso talvez seja o de maior sucesso da atribuição de autoria forense. Outro exemplo seria o caso apresentado por McMenamin (2002, pp. 181-205): uma garota de seis anos de idade, Jon Benét Ramsey, desapareceu, e algum tempo depois foi descoberta morta no porão de sua casa, após terem encontrado uma carta anônima de resgate dentro da residência. Os pais da garota foram considerados suspeitos, e McMenamin ficou encarregado de analisar o bilhete de resgate, comparando-o com outros textos do casal (bilhetes pessoais, cartas, etc.). Sua análise indicou que não era provável que nenhum dos dois suspeitos tivesse escrito o bilhete.

Um caso brasileiro de que se teve conhecimento é o relatado por Harkot-de-La-Taille (2008), em que se analisaram peças de um processo envolvendo a ex-primeira dama de São Paulo, Niceia Pitta. Niceia havia acusado o empresário Jorge Ynes de tê-la ameaçado de morte. A análise de Harkot-de-La-Taille comparou a carta de ameaça com um diálogo (uma entrevista dada à TV Record, no programa “Fala que eu te escuto”) entre o suspeito e um apresentador e mostrou que os enunciadores dos dois textos eram distintos.

Muitos são os métodos existentes para determinar a autoria de texto. A seção 1 apresenta um panorama e as principais características comumente propostas por esses métodos como relevantes para a atribuição de autoria. Quaisquer que sejam os métodos empregados, eles se baseiam na noção de estilo. O estilo pode ser tanto expressão de uma individualidade quanto de uma identidade de grupo. Este trabalho concentra-se na primeira noção. O que se entende por estilo pode ser visto no decorrer das seções 1 e 2.

Uma questão que pode se tornar um problema em casos forenses de atribuição de autoria é o fato de que há variação linguística em um mesmo indivíduo. Se, conforme vem afirmando a Sociolinguística ao longo dos anos, a mesma pessoa emprega formas linguísticas diferentes em situações de comunicação diferentes (e.g. LABOV, 1966[2006]; BELL, 1984; SCHILLING-ESTES, 2001; ECKERT; RICKFORD, 2001) e se esses princípios podem ser aplicados à língua escrita (e.g. McMENANIM, 2002; 2010; OLSSON, 2008; LILLIS, 2013), como determinar que alguém é o autor de vários textos redigidos em situações diferentes e destinados a interlocutores diferentes, como ocorre em muitos, senão na maioria, dos casos de autoria textual no âmbito forense? Essa questão será melhor explorada na seção 2, em que também levanta-se uma hipótese de como esse problema poderia ser resolvido (por meio da análise do plano do conteúdo, conforme a Semiótica Discursiva). A seção 3 conta com um exemplo de análise qualitativa de acordo com essa proposta.

## 1 ANÁLISE DE AUTORIA NO ÂMBITO FORENSE: UM PANORAMA

A análise linguística pode ajudar a determinar a autoria de textos como cartas de suicídio, cartas e mensagens anônimas (por exemplo, em casos de sequestros, chantagens, ameaças, mensagens via celular), *websites* com conteúdo ilegal, *e-mails* anônimos com conteúdo ofensivo, entre outros<sup>7</sup>.

Muito se fala da análise grafotécnica como ferramenta para determinar a autoria textual no âmbito forense. No entanto, uma vez que hoje em dia grande parte dos textos que circulam não são manuscritos, esse tipo de análise não é sempre possível. Juola (2006, p. 2, tradução nossa) afirma: “com o advento da tecnologia moderna da computação, uma quantidade substancial de ‘escritos’ hoje em dia nunca envolve caneta, tinta ou papel”<sup>8</sup>. Em outras palavras, muitas vezes, não há vestígios físicos de autoria para serem examinados. Entra em cena, então, uma análise que busque variantes textuais/discursivas que possam vir a distinguir autores.

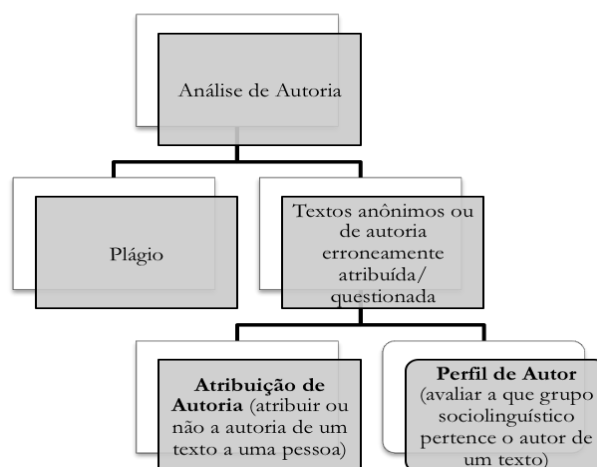
Cabe mencionar que as análises de autoria normalmente se enquadram em duas grandes categorias. A primeira diz respeito a quando há a suspeita de que

---

<sup>7</sup> Diante de textos divulgados pela Internet, pode surgir a seguinte questão: “Uma análise no âmbito da computação (redes) não seria capaz de rastrear o computador fonte de tais textos?”. A resposta é sim, mas esse é o limite de tal análise, pois, como afirma Juola (2006, p. 2), “uma investigação da rede de computadores vai no máximo revelar o computador específico em que o documento foi produzido. É quase impossível descobrir quem estava ao teclado, quem o escreveu” (“A network investigation will at best only reveal the specific computer on which the document was written. It is almost impossible to figure out who was at the keyboard — who wrote it”).

<sup>8</sup> Texto original: “With the advent of modern computer technology, a substantial amount of “writing” today never involves pen, ink, or paper” (JUOLA, 2006, p. 2).

alguém fez passar por suas as palavras de outrem, dito de outro modo, diz respeito aos casos de plágio. A segunda se ocupa de casos em que textos sejam anônimos, assinados por pseudônimos, ou com autoria questionada ou erroneamente atribuída. A segunda categoria pode ainda ser subdividida em duas tarefas – que são frequentemente apresentadas separadamente na literatura sobre o tema, mas que podem ser utilizadas em conjunto em muitos casos – a saber: 1) a comparação entre um texto ou grupo de textos cuja autoria esteja sendo questionada e uma amostra de outros textos de autores suspeitos; 2) a elaboração de um perfil de um autor, mais precisamente, um “perfil sociolinguístico”, já que certos usos linguísticos podem dar pistas de pertencimento a certos grupos sociais. Este trabalho concentra-se na tarefa 1. A Figura 1 ilustra os tipos de análise de autoria (baseado em KREDENS, 2012):



**Figura 1:** Tipos de análises de autoria.

## 1.1 IDIOLETO E ESTILO

A ideia de que é possível saber se alguém é o autor de um texto ou um conjunto de textos baseia-se no pressuposto de que o uso linguístico torna-se, ao longo do tempo, um hábito, assim como se tornam habituais outros comportamentos humanos. Segundo Tomblin (2012, p. 177), existem evidências da Psicolinguística, da Sociolinguística, da Linguística de *Corpus* e dos estudos de Aquisição de Linguagem que mostram que a comunicação é baseada em padrões e em “maneiras preferidas” para expressar as ideias.

Os “hábitos linguísticos” dos indivíduos resultam em seus “idioletos”, ou seja, dois ou mais indivíduos diferentes possuem versões diferentes da língua que falam ou escrevem. Para Coulthard e Johnson (2007, p. 161), o linguista aborda o problema de autoria questionada a partir desse ponto de vista teórico. Ainda que a ideia de que

cada indivíduo tenha um modo particular de usar a língua esteja presente desde quando se apresentou a noção de *ethos*, na *Arte Retórica*, de Aristóteles, o termo “idioleto” só foi cunhado em 1948 por Bloch (KREDENS, 2002; TURELL, 2010; COULTHARD; JOHNSON, 2007, p. 161), para quem ele seria “a totalidade das possíveis elocuições de um falante em um determinado momento enquanto usa a língua para interagir com outro falante”<sup>9</sup> (BLOCH, 1948, p. 7, tradução nossa).

Com o tempo, passa-se a entender o idioleto não como simplesmente “a língua de um indivíduo”, mas sim na relação que essa língua “individual” contrai com a língua “social”. Já no século XX, Labov (1966[2006], p. 5) propõe: “a língua dos indivíduos não pode ser entendida sem conhecimento da comunidade da qual eles fazem parte”<sup>10</sup>. Mais especificamente no âmbito da Linguística Forense, Turell (2010) e McMenemy (2002; 2010) têm posições semelhantes. Turell (2010) cunha o termo “estilo idioleto” como não propriamente o “sistema linguístico de um indivíduo”, mas sim a forma como um sistema linguístico compartilhado por um grupo é usado de modo distintivo por um indivíduo. Para McMenemy (2010, p. 488, tradução nossa), “o estilo linguístico é o resultado do que um indivíduo seleciona para seu uso do conjunto de ferramentas linguísticas disponíveis para o grupo ao qual pertence”<sup>11</sup>.

Pode-se fazer uma correspondência entre os conceitos de idioleto e estilo, na medida em que tanto um quanto o outro dizem respeito ao conjunto de elementos linguísticos que caracterizam um indivíduo. Talvez uma diferença crucial entre eles esteja na abrangência do segundo, conforme previsto por estudos mais recentes da Sociolinguística Variacionista, que enxergam o estilo tanto como caracterizador de um indivíduo quanto indicador de uma afiliação social (IRVINE, 2001; ECKERT; RICKFORD, 2001; COUPLAND, 2007). Tal ideia encontra respaldo em Sapir (1927, p. 900, grifo nosso, tradução nossa), para quem o estilo linguístico é “(...) uma faceta cotidiana da fala *que caracteriza tanto o grupo social quanto o indivíduo*”<sup>12</sup>. Note-se que Sapir já teorizava o estilo como individual ou social e, relativamente ao “estilo individual” – como se conhece hoje, o “idioleto” –, para ele “Nós todos temos nossos estilos individuais [...]. Sempre há um método individual de combinar as palavras em grupos e esses grupos em unidades maiores” (SAPIR, 1927, p. 900, tradução nossa<sup>13</sup>).

---

<sup>9</sup> Texto original: “The totality of the possible utterances of one speaker at one time in using a language to interact with one other speaker” (BLOCH, 1948, p. 7).

<sup>10</sup> Texto original: “the language of individuals cannot be understood without knowledge of the community of which they are members” (LABOV, 1966[2006], p. 5).

<sup>11</sup> Texto original: “linguistic style is the result of what an individual selects to use from the array of linguistic tools available to his or her own group” (MCMENAMY, 2010, p. 488).

<sup>12</sup> Texto original: “an everyday facet of speech that characterizes both the social group and the individual” (SAPIR, 1927, p. 900).

<sup>13</sup> Texto original: “We all have our individual styles [...]. There is always an individual method (...) of

A ideia de estilo ou “estilo idioletal” presente na literatura sobre atribuição de autoria é a de “cosseleção” ou “combinação” de características linguísticas. De acordo com Rudman (1998, p. 360), as propriedades linguísticas comumente indicadas como potenciais marcadores de estilo (por exemplo, tamanho de palavras, tamanho de sentenças, relação “type/token”) não podem ser utilizadas isoladamente, mas se tornam importantes se consideradas em conjunto com outros aspectos. Conforme Love (2002, p. 101), não é apenas através de uma ou outra característica que se pode reconhecer um estilo, mas pela repetição, pela consistência de uma combinação de características. Para McMenamin (2002, p. 51), o estilo linguístico é dado pela recorrência de uma combinação de traços linguísticos, uma “constelação de variáveis”, e não pelo estabelecimento de um único traço isoladamente, e isso se deve ao fato de que formas linguísticas únicas, singulares, são raras. Na mesma linha, segue Coulthard (2006, p. 1, tradução nossa), segundo quem: “embora a princípio qualquer falante ou escritor possa usar qualquer palavra a qualquer momento, eles tendem a usar tipicamente a mesma cosseleção de palavras”<sup>14</sup>. Assim, a identificação de um autor nunca será feita com base em apenas uma característica linguística, mas na combinação de várias dessas. Tais características podem aparecer também isoladamente em outros autores, mas é improvável que o mesmo conjunto apareça igualmente em outro autor (McMENAMIN, 2002, p. 172).

Em termos de filiação teórica, vê-se que as duas principais noções da análise de autoria – idioleto e estilo (este último, individual ou social) – bebem da fonte dos estudos sociolinguísticos iniciados por Labov (1966[2006]), na medida em que, assim como a Sociolinguística, preocupam-se com a relação entre a variação intrafalante e interfalante, ou seja, entra a variação existente em um mesmo indivíduo e em um grupo de indivíduos.

Há muitos casos em que o estilo social pode ser utilizado em análises de autoria, mas ele se relaciona à segunda tarefa mencionada anteriormente – a elaboração de perfis sociolinguísticos – que não é foco deste trabalho que, portanto, não se alonga em sua descrição.

No senso comum, seja em termos linguísticos ou considerando qualquer outra manifestação (por exemplo, vestuário, decoração), impera uma oposição entre “ter” ou “não ter” estilo. O indivíduo “estiloso” é aquele que faz uso de algum recurso que se destaca, aquele cujas escolhas diferem sobremaneira das escolhas da maioria. Essa maneira de pensar o estilo vai ao encontro tanto da ideia de estilo como

---

arranging words into groups and of working these up into larger units” (SAPIR, 1927, p. 900).

<sup>14</sup> Texto original: “whereas in principle any speaker/writer can use any word at any time, in fact they tend to make typical and individuating co-selections of preferred words” (COULTHARD, 2006, p. 1).

expressividade quanto como desvio, a primeira proveniente de uma Estilística Linguística, filiada a Bally (1941; 1952 apud FIORIN, 2008, p. 93) e a segunda advinda de uma Estilística Literária filiada a Spitzer e Damaso Alonso (FIORIN, 2008, p. 95). Ainda no senso comum, outra concepção de estilo é aquela ligada a um conjunto de características (por exemplo, tendências, gostos, comportamentos, modos de falar) que podem ser associadas tanto a um indivíduo quanto a um grupo. Essas duas concepções têm em comum o fato de considerar o estilo como algo distintivo. O estilo, na língua ou em outros sistemas semióticos, tem a ver com distinção, conforme observa Irvine (2001, p. 22, tradução nossa<sup>15</sup>): “O que quer que ‘estilos’ sejam, em linguagem ou em qualquer outra coisa, eles são parte de um sistema de distinção em que um estilo contrasta com outros estilos possíveis e o sentido social que estes estilos veiculam também contrastam entre si”.

Mesmo que as duas grandes vertentes sobre o estilo mencionadas aqui tenham como base a mesma ideia de distinção, para fins de atribuição de autoria, a ideia de estilo como expressividade e a oposição entre “com estilo” e “sem estilo” é ignorada e prevalece a ideia de que todo autor tem um estilo. Isso é importante principalmente em se tratando do contexto forense, em que, como visto anteriormente, figuram textos que muitas vezes não foram elaborados por profissionais da escrita. Se o sentido que permeia as várias definições de estilo está relacionado à distinção ou à diferença e se todo autor tem um estilo, seu estudo é fundamental quando se trata de identificar a autoria de um texto ou grupo de textos, ou de excluir suspeitos da lista de possíveis autores. Para uma análise do tipo que é proposta aqui, o estilo como concebido pela Semiótica Discursiva ganha destaque. Adiante, a seção 1.4 tratará do conceito de estilo neste quadro teórico.

## **1.2 TEXTOS FORENSES E NÃO FORENSES**

Os textos sujeitos à análise de autoria na esfera forense podem ser de dois tipos: alguns são claramente “criminosos”, tais como cartas de ameaças, bilhetes de resgate, conversas na internet sobre sexo envolvendo menores, cartas de suicídio, etc.; outros são mais “cotidianos” e não inerentemente criminais, como diários, e-mails pessoais, cartas. Estes últimos podem ser utilizados como base para análises comparativas ou para fornecer um alibi em uma investigação (GRANT, 2008, p. 216). Grant (2008, p. 216) argumenta que, devido à variedade de tipos de texto que

---

<sup>15</sup> Texto original: “Whatever ‘styles’ are, in language or elsewhere, they are part of a system of distinction, in which a style contrast with possible styles, and the social meaning signified by the style contrasts with other social meanings” (IRVINE, 2001, p. 22).



ocorrem no contexto forense, generalizações sobre as suas características devem ser cautelosas. Apesar dessa variedade, é possível observar alguns aspectos que distinguem esses textos daqueles tipicamente analisados em investigações de autoria no âmbito literário, histórico ou religioso. Estes últimos textos são chamados de “não forenses”. Uma das principais diferenças percebidas é que os textos “não forenses” são mais bem elaborados. Nas palavras de Grant (2008, p. 216), “o autor provavelmente gastou um tempo considerável pensando no texto e compondo-o”.

Outra consideração é que os textos não forenses são, via de regra, escritos por profissionais ou pelo menos por pessoas com um nível alto de educação formal. Ao contrário, os textos forenses são mais espontâneos, escritos, na maioria das vezes, por “naïve writers”<sup>16</sup> (BAYEEN *et al.*, 2002)<sup>17</sup>. Na atribuição de autoria não forense, há, frequentemente, grandes quantidades de textos a serem analisados, enquanto nos contextos forenses, isso quase nunca ocorre, e os textos são, em geral, bastante curtos. O Quadro 1 resume as principais diferenças entre as duas classes de textos:

<b>Textos “forenses”</b>	<b>Textos “não forenses” (literários, religiosos, históricos, etc.)</b>
Curtos	Longos
Espontâneos/não planejados	Não espontâneos/planejados
Direcionados a uma pequena audiência	Direcionados a grandes audiências
Autores não são profissionais	Autores geralmente “profissionais da escrita” ou com nível de escolaridade maior

**Quadro 1:** Diferenças entre textos forenses e não forenses.  
Fonte: Grant (2008, pp. 216-217); Turell (2012, *slide 8*).

### 1.3 PANORAMA

Como já mencionado, a tarefa de identificar o autor de um texto pode surgir em outros contextos, além do judicial. Estudos de textos literários, históricos ou religiosos estão entre aqueles que enfocam essa questão. Além desses campos, há situações mais cotidianas que poderiam suscitar a pergunta “Quem escreveu este texto?”, como, por exemplo, conforme indica Grant (2004, p. 17), em assuntos políticos atuais ou na área de Educação. Disso decorre que há muitos trabalhos publicados sobre o tema, derivados tanto da pesquisa de estudiosos da linguagem quanto de outros ligados à computação. Esse panorama, porém, dá mais ênfase às pesquisas que possam ser aplicadas especificamente no contexto forense.

Encontram-se abordagens diferentes para o problema de identificação de

<sup>16</sup> “Autores ‘ingênuos’”.

<sup>17</sup> Os textos analisados na seção 3 podem ser categorizados como “forenses”, de acordo com as características elencadas no Quadro 1. Aquelas características orientaram a coleta de dados para o *corpus* de nossa pesquisa.

autoria e, segundo Grant (2004, p. 43, tradução nossa), isso “parece refletir diferentes hipóteses sobre a natureza da linguagem e a linguística”<sup>18</sup>. Vários dos trabalhos existentes apoiam-se em métodos quantitativos computacionais, no paradigma conhecido como Estilometria, tendo como base, por exemplo, a frequência lexical de palavras ou expressões, o comprimento das sentenças e o comprimento de palavras. São exemplos desses trabalhos Winter e Woolls (1996) e Baayen *et al.* (2002). Chaski (1997; 2001; 2004), por sua vez, concentra-se na análise sintática e no uso da pontuação como medida de diferenciação entre autores. Outras pesquisas, numa vertente considerada mais qualitativa, apoiam-se, principalmente, na observação da variação linguística, como é o caso da Estilística Forense (McMENANIN, 2002), da noção de “estilo idioletal” (TURELL, 2010; 2012), e dos estudos de Olsson (2004). Há ainda trabalhos mais voltados para uma análise de vocabulário (WOOLLS; COULTHARD, 1998; COULTHARD, 2004b) e a chamada “*Forensic discourse analysis*” – análise do discurso forense – Coulthard (1992; 1993; 1994; 1999; 2001 apud COULTHARD, 2006).

Da mesma maneira que a divisão anterior, Grant (2013, p. 470) propõe que as análises de autoria podem ser separadas em duas grandes vertentes: a Estilometria e a Estilística, sendo a primeira a que procura descobrir marcadores quantificáveis de autoria, envolvendo a medição das similaridades e das diferenças entre autores diferentes (HOLMES, 1998; GRANT, 2004); a segunda é a que engloba todas as outras análises em que a quantificação de características não seja o ponto principal. O uso de uma ou outra abordagem também pode depender do tipo de texto que se quer analisar. A Estilometria é bem-sucedida em casos em que haja “textos mais longos que envolvem algumas milhares de palavras” (GRANT, 2013, p. 471, tradução nossa)<sup>19</sup>. Já quando a análise encontra textos curtos ou fragmentários (o que é mais comum ocorrer no contexto forense), os analistas têm empregado a Estilística Forense e não a Estilometria. Nas abordagens estilísticas, não se propõe que apenas um ou outro marcador seja utilizado sempre, independentemente do caso, como um universal, mas sim que cada indivíduo apresenta um conjunto de características que o identifica, e este conjunto pode variar entre indivíduos. Nas palavras de Coulthard e Johnson (2007, p. 173, tradução nossa), “não é raro que um especialista utilize mais de uma abordagem”<sup>20</sup>.

Este trabalho se alinha à vertente estilística para análise de autoria, porém,

---

<sup>18</sup> Texto original: “[they] appear to reflect different assumptions about the nature of language and linguistics” (GRANT, 2004, p. 43).

<sup>19</sup> Texto original: “longer texts involving some thousands of words” (GRANT, 2013, p. 471).

<sup>20</sup> Texto original: “It is not unusual for the expert to use more than one approach” (COULTHARD; JOHNSON, 2007, p. 173).

apresenta-se a seguir um breve percurso dos estudos estilométricos, já que eles compreendem a base, o início dos estudos de autoria, forense ou não. Baayen *et al.* (2002) introduzem a área de maneira sucinta, a ver:

As tentativas estilométricas de rastrear a autoria de textos de autores desconhecidos ou contestados vêm de longa data. Elas têm sido aplicadas a textos importantes, tais como a Bíblia, as obras de Shakespeare, e os *Federalist Papers*. Uma grande variedade de técnicas provenientes de diversas disciplinas têm sido consideradas, desde análises estatísticas multivariadas até redes neurais e aprendizado de máquinas. Muitos aspectos textuais diferentes têm sido analisados, desde o comprimento das palavras ou sentenças até as palavras ou características linguísticas mais comuns ou ou mais raras (BAAYEN *et al.*, 2002, p. 1, tradução nossa).<sup>21</sup>

A Estilometria não seria um método propriamente dito, mas um paradigma que abriga diversos trabalhos de identificação de autoria (não apenas em contextos forenses) que têm em comum a análise estatística, com a quantificação de características linguísticas, frequentemente com auxílio computacional (GRANT, 2004). O primeiro estudo que procurou identificar um marcador de autoria foi o de Augustus de Morgan, em 1851, cujo objeto eram algumas cartas atribuídas a São Paulo. Sua hipótese era de que o tamanho das palavras usadas por diferentes autores seria um marcador (HOLMES, 1998; COULTHARD; JOHNSON, 2007). Em 1887, Mendenhall testou tal hipótese no problema de autoria das epístolas paulinas (GRANT, 2004, p. 20; COULTHARD; JOHNSON, 2007, p. 164), contando manualmente o tamanho das palavras nas cartas, comparando essa medida com aquela de mais de trezentas mil palavras provenientes de escritos bíblicos, de trabalhos de Bacon, Marlowe e Shakespeare. Seu trabalho sugeriu que havia semelhança entre as epístolas verificadas e os escritos de Marlowe e Shakespeare. A principal característica do método de Mendenhall, isto é, isolar uma característica linguística como marcador de autoria, foi bastante repetida em estudos estilométricos. Diz-se que esse trabalho deu início ao que se conhece hoje por Estilometria.

Nas décadas que seguiram o trabalho de Mendehall, pouco foi feito em estilometria até 1932 quando Zipf publicou seu trabalho sobre a frequência lexical nos textos, o que ficou conhecido como “Zipf’s Law” (Lei de Zipf). De acordo com Grant (2004, p. 23), “tais medidas de frequência lexical [...] encorajaram a pesquisa

---

<sup>21</sup> Texto original: “Stylometric attempts to trace the authorship of texts by unknown or contested authors have a long history. They have been applied to influential texts such as the Bible, the works of Shakespeare, and the Federalist Papers. A wide variety of techniques from many disciplines have been considered, from multivariate statistical analysis to neural networks and machine learning. Many different facets of texts have been analysed, from sentence and word length to the most common or the rarest words, or linguistic features” (BAAYEN *et al.*, 2002, p. 1).

de vários outros métodos e medidas estilométricas ao longo dos anos 1940, 1950 e 1960, o que culminou no exame da autoria dos *Federalist Papers*, conduzido por Mosteller e Wallace (1963)”.

*The Federalist Papers* são tratados políticos publicados em 1787 e 1788 nos Estados Unidos durante o que ficou conhecido como o “*constitutional referendum*” do estado de Nova Iorque (85 textos publicados anonimamente). Atualmente, sabe-se que são três os seus autores: Alexander Hamilton, John Jay e James Madison. Entretanto, há 12 desses textos cuja autoria foi disputada por Madison e Hamilton. Mosteller e Wallace conduziram uma investigação linguística que procurou diferenças idioletais em textos conhecidos desses dois autores, comparando-as com os 12 textos questionados. Seu estudo concluiu que todos os 12 textos pertenciam a Madison, o que encontrou respaldo em outras pesquisas feitas por historiadores (COULTHARD; JOHNSON, 2007; GRANT, 2004).

Os anos 1980 assistiram a uma virada nos estudos estilométricos, representada principalmente pelos trabalhos de John Burrows (GRANT, 2004) e, já nos anos 1990, Holmes (1992). Esses trabalhos passaram a fazer uso de *corpora* sintaticamente anotados e de análises estatísticas e computacionais diferentes das realizadas até o momento, tais como as cadeias de Markov, a *Principal Component Analysis*, e abordagens relacionadas ao aprendizado de máquinas (CORNEY, 2003).

A década de 1990 foi praticamente dominada pelas discussões acerca do método CUSUM (*Cumulative Sum*), de Andrew Morton (GRANT, 2004; COULTHARD; JOHNSON, 2007). Um resumo do método, sobre o qual houve muitas controvérsias, é apresentado por Coulthard e Johnson (2007, p. 167). Basicamente, o método levava em conta não o texto, mas a sentença como unidade básica e calculava a frequência com que apareciam substantivos, palavras começadas com vogais ou palavras compostas com duas, três ou quatro letras.

Outro marcador proposto nesta época foi o tamanho médio das sentenças combinado com a “riqueza” lexical (WINTER; WOOLLS, 1996). Entra-se aí em estudos que passaram a ser conhecidos como “Análise de vocabulário”. Tais estudos têm suas raízes na proposição de Honoré (1979), que postulou que a frequência de *hapax legomena* (isto é, palavras que aparecem uma única vez no texto) seria um modo eficaz de medição do vocabulário. O problema com a técnica de Honoré estava no fato de não se distinguir entre itens gramaticais (muito frequentes) e lexicais. Assim, Winter e Woolls (1996) resolveram o problema excluindo as palavras gramaticais da medição. Resultados sobre estudos realizados com essas medidas podem ser verificados em Coulthard e Johnson (2007, p. 165).

Mais recentemente, Kredens (2000) testou o potencial discriminatório de marcadores discursivos em textos orais, comparando entrevistas de dois músicos britânicos, Robert Smith e Steven Morrissey. Seus resultados indicaram 3 marcadores discursivos como significativos (*like, I mean, kind of/sort of*) para distinguir os seus idioletos. Chaski (1997; 2001; 2005) propõe que a pontuação, mais precisamente a função sintática da pontuação, seja um marcador relevante.

Sobre a abordagem conhecida como *Forensic Discourse Analysis*, sua principal tarefa é identificar usos linguísticos que estejam “fora de lugar” (GRANT, 2004, p. 43). Por exemplo, ao analisar o registro escrito de um depoimento, pode-se ter uma “dica” de que ele foi forjado quando se encontram marcas do discurso policial no que deveria ser apenas a transcrição da fala de um “cidadão comum”.

Saindo da vertente estilométrica, diversos trabalhos têm insistido em apontar não apenas um elemento linguístico como distintivo, como faz a Estilometria, mas uma combinação deles na busca por um modo eficaz de se determinar a autoria de um texto num “contexto” comparativo (e.g. TURELL, 2010; McMENANIN, 2002). Descrever as diferentes “constelações” de variáveis e variantes empregadas por um determinado autor é a proposta de McMenanin (2002; 2010), chamada por ele de “Estilística Forense”. McMenanin parte da Sociolinguística Variacionista, cujo pressuposto é o de que os usos linguísticos são inerentemente variáveis. De acordo com Labov (1966[2006], p. 3), “a variabilidade é parte integrante do sistema linguístico”. Ela é um fato observável tanto no uso linguístico de pessoas diferentes, dentro de uma mesma comunidade, quanto nos usos que os mesmos indivíduos fazem da língua nas diferentes situações comunicativas. Uma variável é um conjunto de duas ou mais formas linguísticas de se dizer “a mesma coisa” (LABOV, 1972). Para McMenamim, o conjunto de variáveis e variantes recorrentes podem indicar a autoria.

Em português, trabalhos sobre autoria têm sido desenvolvidos principalmente por Silva (e.g. SILVA, 2008; SILVA *et al.*, 2010). Silva (2008, p. 95) verifica a “utilidade e aplicabilidade” da elevada frequência de palavras e expressões como marcador. Em tal estudo, analisou-se a frequência dos tetagramas (combinação de quatro palavras) utilizados uma única vez (*hapax legomena*) e “a frequência de expressões que ocorrem mais vezes nos textos do mesmo autor (*hapax dislegomena*)” (SILVA, 2008, p. 95). Brandão (2006) é mais um exemplo de trabalho em atribuição de autoria. O autor propôs a utilização de um método que tem como principal característica a contagem dos itens funcionais presentes nos textos. Em seu trabalho de 2006, analisaram-se as *Cartas Chilenas* – um conjunto formado por quatorze

poesias do século XVIII, atribuídas a Tomás Antônio Gonzaga, assinadas, porém, com o pseudônimo “Critilo”.

Alguns trabalhos propõem que os elementos linguísticos encontrados, principalmente os idiossincráticos, sejam comparados com um *corpus* de referência para que assim se possa atestar sua raridade (COULTHARD, 1994; TURELL, 2010).

Turell (2010) também utiliza como característica linguística a sequência de categorias (classes) linguísticas, os erros ortográficos e a utilização de construções que não estejam de acordo com a norma padrão (TURELL, 2010). Sobre a sequência de categorias gramaticais como distintivas de um autor, Turell (2010, p. 232) cita ainda uma série de estudos que também se basearam nessa característica.

Uma relação das características normalmente observadas nos trabalhos aqui mencionados pode ser vista a seguir, no Quadro 2:

<b>Elementos frequentemente verificados em análises de autoria</b>
- Relação “ <i>Type-Token</i> ”
- Frequência lexical
- Distribuição e frequência de classes de palavras
- Diversas variáveis morfológicas e sintáticas
- Tamanho e distribuição de palavras (em letras ou em sílabas), ou de sentenças, ou de parágrafos; sequência de letras
- Distribuição dos N-gramas
- Pontuação
- Uso de palavras ou “ <i>collocations</i> ” raras
- Ortografia

**Quadro 2:** Elementos frequentemente verificados em análises de autoria.

Os métodos existentes para atribuição de autoria podem, ainda, ser classificados em “qualitativos” ou “quantitativos”, de acordo com a metodologia empregada, embora uma abordagem mista seja, na prática, bastante utilizada. Ao longo da literatura sobre análise de autoria, é comum encontrar uma divisão dos estudos existentes – mais metodológica do que paradigmática – em “qualitativos” ou “quantitativos”. Numa análise qualitativa, identificam-se e descrevem-se elementos característicos de um autor. Já numa análise quantitativa, separam-se e medem-se de alguma forma esses elementos.

Muitas das análises mais recentes combinam essas duas abordagens. Segundo Gibbons (2011, p. 249), “as variáveis selecionadas para análises quantitativo-estatísticas ou computacionais são normalmente provenientes de conhecimento qualitativo por parte do especialista [...]”<sup>22</sup>. Dessa forma, melhor do que ver essas duas abordagens como diferentes métodos de análises é considerá-las como passos

---

<sup>22</sup> Texto original: “The variables selected for quantitative statistical or machine analysis are usually derived from the qualitative knowledge and understanding of the expert [...]” (GIBBONS, 2011, p. 249).

da análise, já que a análise qualitativa é geralmente necessária para identificar as similaridades e diferenças entre os textos, mas pode não ter caráter probatório – daí entra a parte quantitativa da análise (GIBBONS, 2011, pp. 249-250). Turell (2010) e McMnamin (2002) também reconhecem que esses dois tipos de análise são complementares.

Por meio da análise qualitativa buscam-se semelhanças e diferenças entre o texto questionado e o conjunto de textos conhecidos. De acordo com McMnamin (2002), os resultados qualitativos são mais fáceis de demonstrar em julgamentos. Segundo ele, essas evidências apelam para o senso de probabilidade não matemático das pessoas. Outro motivo para a apresentação de análises qualitativas tem relação com o fato de a quantidade de textos para analisar ser normalmente pequena nos contextos forenses, muitas vezes impossibilitando o fornecimento de evidências quantitativas. Segundo Olsson (2008, p. 7, tradução nossa<sup>23</sup>), “grande parte dos trabalhos em Linguística Forense é de observação e análise em oposição a medidas de probabilidade”.

Junto ao aumento do interesse na análise de autoria no âmbito forense, surge uma dúvida sobre quão “científica” ela pode ser. De acordo com Grant (2007, p. 2), “científico” nas discussões em torno da questão geralmente está ligado à quantificação ou ao tratamento estatístico dos dados, principalmente nos Estados Unidos, onde há regras, conhecidas como “*the Daubert criteria*”, que regulam a apresentação de evidências técnicas ou científicas em um tribunal<sup>24</sup>. Desses critérios, o terceiro é o que está relacionado mais diretamente a uma quantificação dos dados. Ao contrário dos Estados Unidos, o Reino Unido confia mais na expertise do perito do que no método (COULTHARD, 2012).

Para McMnamin (2002), apesar das limitações relacionadas às análises quantitativas de autoria, elas são muito relevantes, na medida em que métodos quantitativos podem tornar os testes de hipóteses mais confiáveis e mais fáceis de se fazer e atender aos requisitos externos por evidências científicas.

Qualquer que seja a metodologia de análise, os métodos para atribuição de autoria, apesar de voltarem-se para características linguísticas diferentes, têm em comum o fato de sempre tentarem identificar características que sejam ou não próprias de um autor. Tais elementos podem ser encontrados nos diferentes níveis

---

<sup>23</sup> Texto original: “a good deal of forensic linguistic work is observation and analysis as opposed to the measurement of probability” (OLSSON, 2008, p. 7).

<sup>24</sup> De acordo com “*the Daubert Criteria*”, os critérios a que evidências apresentadas pelos peritos devem atender são: a) A teoria foi testada? b) O método ou a teoria foi avaliada por outras pessoas da área (*peer reviewed*)? c) Há uma taxa de erro conhecida? d) O método é geralmente aceito? (GRANT, 2007; SOLAN; TIERSMA, 2005).

linguísticos dos textos, ou seja, podem estar tanto no léxico quanto no nível ortográfico, ou morfológico, ou sintático, conforme foi visto nesta seção.

#### 1.4 SEMIÓTICA DISCURSIVA E ESTILO

Herdeira da Linguística saussuriana, a Semiótica Discursiva é uma teoria da significação que busca desvendar os mecanismos de construção do sentido nos diversos tipos de texto. Além de ser uma teoria da significação, a Semiótica é também uma metodologia de análise de textos (ALMEIDA, 2009, p. 11).

Na esteira de Hjelmslev (1943[2003], p. 53), a Semiótica baseia-se na ideia de que um texto é um signo e que este, por sua vez, é “um todo formado por uma expressão e um conteúdo”. Um conteúdo é, sempre, manifestado por uma expressão. O plano do conteúdo é abordado pelos semioticistas sob a forma de um “percurso gerativo do sentido” (GREIMAS; COURTÉS, 1979[2008]), um simulacro metodológico de uma estrutura analisável, uma “gramática” que está presente em todos os textos. Desde o nível mais profundo do percurso gerativo até a manifestação textual, passa-se de estruturas mais simples e abstratas (pressupostas) para mais complexas e concretas (pressupostas). A análise semiótica vai para além da manifestação dos textos e para além da aparência, buscando os seus sentidos imanentes. A Figura 2 representa tal modelo:



**Figura 2:** Semiótica Discursiva: plano da expressão e plano do conteúdo com o percurso gerativo do sentido.

Como se verifica na Figura 2, são três as etapas do percurso gerativo do sentido, que são resumidas a seguir.

1) *Nível fundamental:* nível mais abstrato que apresenta uma oposição semântica mínima que permeará todo o texto e estará na base de sua construção. Em desenvolvimentos mais recentes, no âmbito da Semiótica Tensiva (TATIT, 2001; ZILBERBERG, 2006a; 2006c), o nível fundamental tem sido visto como um nível em que há oposições ainda mais abstratas do que “categorias semânticas”, da ordem, por



exemplo, de continuidades e descontinuidades, ou relaxamento e tensão. Fala-se, então, em um “nível tensivo”.

2) *Nível narrativo*: organização narrativa que está “por trás” de todos os textos; costuma ser formulada em termos de transformações, envolvendo o ponto de vista de um “sujeito” que está em busca de um “objeto”.

3) *Nível discursivo*: o revestimento concreto dos níveis anteriores mais abstratos, levando-se em conta os graus de figuratividade dos conteúdos propostos, bem como as múltiplas estratégias de que se vale o enunciador para engendrar efeitos de proximidade ou distância enunciativa.

Ainda no plano do conteúdo, é possível reconhecer um componente modal e um aspectual nos textos. Apesar de esses componentes serem tradicionalmente incorporados na análise semiótica, sua reformulação como “níveis” do percurso gerativo só foi feita mais recentemente por Zilberberg (2006b, p. 156), que os chama de nível aspectivo e nível modal.

Com tal importância para o plano do conteúdo dentro da teoria semiótica, é de se esperar que para ela o estilo considere também o que acontece nesse nível. O estilo é definido como “um conjunto global de traços recorrentes do plano do conteúdo (formas discursivas) e do plano da expressão (formas textuais), que produzem um efeito de sentido de identidade. Configuram um *ethos* discursivo, ou seja, uma imagem do enunciador” (FIORIN, 2008, p. 97). Estilo é, então, recorrência e distintividade (ou diferenciação) (FIORIN, 2008; DISCINI, 2009). Fiorin (2008, pp. 96-102) explica:

Estilo é o conjunto de traços particulares (...). É o conjunto de características que determina a singularidade de alguma coisa (...). O termo estilo alude, então, a um fato diferencial: diferença de um autor em relação a outro, de um pintor relativamente a outro, de uma época em relação a outra, etc. (...). Diferença e repetição são dois momentos no processo dinâmico de produção estilística. O estilo aparece como diferença e, em seguida, fixa-se em esquemas (...).

A Semiótica postula que as escolhas de um enunciador se dão em todos os níveis, mesmo os mais abstratos. O enunciador seleciona “os valores praticados em todos os estratos gerativos, das estruturas profundas, às de superfície” (TATIT, 2001, p. 19). Sendo assim, o estilo é controlado pela instância da enunciação e, portanto, aparece tanto no plano da expressão quanto no plano do conteúdo. Segundo Discini (2009, p. 28), “cumpre o analista de estilo (re)construir o ator da enunciação de uma totalidade de discursos”. Na atribuição de autoria forense, esse ator da enunciação será relacionado a um indivíduo “de carne e osso”. Conforme explica Harkot-de-La-

Taille (2008, p. 2),

A figura dos laudos periciais acerca da autoria de determinado texto pressupõe a possibilidade de, por meio de uma dada produção textual, estabelecerem-se, com base em marcas linguístico-discursivas, relações de identidade entre o sujeito do discurso – “de papel” – e um cidadão – sujeito “de carne e osso”.

Para Greimas e Courtés (1979[2008], pp. 180, 253), como característica pessoal de um escritor, o estilo está próximo da noção de idioleto. O idioleto está ligado à variação (indivíduos diferentes usam a língua de forma diferente), e o estilo significa que essa variação estará recorrentemente presente na língua de um indivíduo.

Os níveis do plano do conteúdo mencionados anteriormente serão brevemente explicados no momento da análise que se propõe na seção 3. Para uma revisão exhaustiva do campo da Semiótica Discursiva, ver Barros (2001; 2003), Fiorin (2005; 2008), Greimas e Courtés (1979[2008]), Greimas e Fontanille (1993), Zilberberg (2006a; 2006b; 2006c) e Tatit (2001).

## **2 PROBLEMAS E HIPÓTESES**

Conforme se pôde perceber na seção anterior, a literatura sobre atribuição de autoria na esfera forense conta com diversos trabalhos significativos. No entanto, esses trabalhos deixam de apresentar uma solução no que diz respeito à variação intrafalante (ou estilística), isto é, na variação no uso linguístico de um mesmo indivíduo<sup>25</sup>. A hipótese central dos estudos de atribuição de autoria é a de que cada autor tem um estilo (LOVE, 2002, p. 12), mas, no contexto forense, os textos comparados são frequentemente de naturezas distintas. Por exemplo, o texto de autoria questionada pode ser uma carta de suicídio, e os textos coletados para a comparação podem diferir, por exemplo, quanto ao propósito, tema, grau de formalidade, destinatários ou gênero (podem ser mensagens de celular, posts em redes sociais, e-mails formais, cartas comerciais, e-mails para amigos, textos de blogs, diários, entre outros gêneros). Ou seja, não há nenhuma garantia – na verdade é até mesmo improvável – que se consigam textos que foram escritos no mesmo tom informal que a carta de autoria questionada e muito menos provável é que se consigam outras cartas de suicídio do mesmo autor para a comparação. Assim, emerge a pergunta: se, conforme afirma a Sociolinguística, existe variação linguística

---

<sup>25</sup> A questão da variação intrafalante, também chamada de variação estilística, será tratada mais detalhadamente adiante.

num mesmo indivíduo, ou seja, se a mesma pessoa emprega formas linguísticas diferentes em situações de comunicação diferentes (LABOV, 1966[2006]; BELL, 1984; SCHILLING-ESTES, 2001; ECKERT; RICKFORD, 2001) e se esses princípios podem ser aplicados à língua escrita (McMENANIM, 2002; 2010; OLSSON, 2008; LILLIS, 2013), como determinar que alguém é o autor de vários textos redigidos em situações diferentes e destinados a interlocutores diferentes?

Verifica-se que os trabalhos existentes analisam elementos linguísticos observados no que Hjelmslev chamou de “plano da expressão” (HJELMSLEV, 1943[2003]). As características comumente examinadas são a pontuação, a utilização de maiúsculas, o modo de datação, a saudação inicial, o modo de anotar os números, os erros de ortografia, as expressões e palavras de elevada frequência, a forma de escrever as abreviaturas, as variáveis morfológicas, as variáveis sintáticas, o “design” geral das páginas, etc. Entretanto, o linguista dinamarquês propõe que a língua, tal como outros sistemas semióticos, é a junção do “plano da expressão” ao “plano do conteúdo”, dois planos inter-relacionados. Ora, se a língua é tanto expressão quanto conteúdo, por que não olhar para marcadores estilísticos no segundo?

Para Fiorin (2008) e Discini (2009), o estilo é um conjunto de marcas de um modo de dizer, e tais marcas estão presentes desde os níveis mais “enraizados” dos textos e não só na sua manifestação. Assim, a primeira hipótese é que a atribuição de autoria pode ser mais acurada se forem considerados os dois planos da linguagem, em vez de apenas o da expressão.

Quanto ao problema da variação intrafalante, a hipótese é que ela, abundante no plano da expressão, tenda a ser menor conforme se “avança” para os níveis mais “profundos” do plano do conteúdo. Quanto mais abstrato/simples o nível no plano do conteúdo, menos opção de escolha para os falantes, o que resulta em uma chance maior de se escolher sempre a mesma forma de organização do conteúdo. Se, com uma grande chance de fazer sempre as mesmas escolhas, dois indivíduos distintos escolhem recorrentemente as mesmas opções, diferentemente um do outro, isso significa que a opção de cada um tem grande poder discriminatório.

O objetivo aqui é, então, propor um modo de analisar um conjunto de textos em busca de características estilísticas individuais de um autor, de traços menos variáveis, ainda que os textos sob análise sejam de natureza muito diferente, como os mencionados anteriormente.

Outra vantagem que a Linguística Forense pode obter ao analisar os textos desse modo é o fato de que os textos dos casos forenses de identificação de autoria, como já referido, são frequentemente curtos, dificultando o estabelecimento de

características recorrentes do plano da expressão.

Em suma, pode-se dizer que a pesquisa que originou este trabalho aborda questões relevantes para a análise forense de autoria, que se tornam obstáculos quando se observam certos pressupostos provenientes da Sociolinguística Variacionista e da Semiótica Discursiva, a saber:

a) Por que considerar apenas elementos linguísticos do plano da expressão nas análises de autoria, se a língua é tanto expressão quanto conteúdo (HJELMSLEV, 1943[2003]) e o conteúdo pode ser estruturado numa gramática (GREIMAS; COURTÉS, 1979[2008])?

b) Como determinar que alguém é o autor de diversos textos quando a Sociolinguística demonstra que existe variação intrafalante, dependente de tantos fatores?

c) Como identificar características recorrentes se os textos que aparecem na análise forense de autoria são frequentemente curtos?

Propõe-se que a resposta esteja na incorporação da análise do plano do conteúdo, conforme a Semiótica Discursiva, às análises de autoria.

### **3 EXEMPLO DE ANÁLISE**

#### **3.1 DO CORPUS**

Para este exercício de análise foi selecionado aleatoriamente um conjunto de textos que compõe um *corpus* maior que está sendo coletado, chamado de *Corpus Geral para Análise de Autoria (CGAA)*. O conjunto conta com 21 textos de um mesmo autor (um adulto, com idade entre 26 e 40 anos, do sexo feminino e com nível de escolaridade superior). São 21 textos, dentre os quais um foi selecionado aleatoriamente, simulando um texto questionado. Vale ressaltar que todos esses textos são “reais”, isto é, não foram escritos “sob demanda”; já haviam sido produzidos *a priori* no momento da sua coleta.

Neste momento, simula-se a tarefa de classificar o texto questionado como pertencente a um grupo de textos. Assim, existe um texto questionado e 20 textos de autoria conhecida (que se sabe serem de um mesmo autor). A pergunta que se quer responder é: “É possível que o texto questionado tenha sido escrito pelo mesmo autor dos textos de autoria conhecida?”.

Os textos coletados desse autor compreendem e-mails profissionais, pessoais, um texto de um blog e, em sua maior parte, posts na rede social Facebook.

Pode-se indagar sobre o fato de os textos aqui utilizados não serem textos de casos forenses reais. Sobre isso, vale ressaltar que Turell (2012) divide o objeto de estudos da Linguística Forense em dois tipos: a) “dados reais” (“real world data”) e b) “dados de casos de Linguística Forense” (“*real Forensic Linguistics case data*”).

Os textos aqui analisados são do primeiro tipo. Observe-se que os textos que fazem parte do *corpus* não foram objeto de disputa judicial, nem de investigação criminal. No entanto, isso não resulta em problema, já que muitos dos métodos conhecidos nesta área foram testados em outro tipo de *corpus*, antes de serem usados em contendas judiciais, e essa opção é até mesmo preferida por alguns autores (GRANT, 2004, p. 22).

Embora os textos sob análise sejam, em sua maioria, muito diferentes quanto ao tema e, algumas vezes, provenientes de diferentes situações de comunicação, a análise do seu nível profundo (plano do conteúdo), de acordo com o modelo da Semiótica Discursiva, indicou que há recorrências que remetem a um mesmo *ethos*. Essas recorrências foram percebidas em diferentes instâncias de organização do plano do conteúdo. As características arroladas a seguir como aquelas relevantes como um marcador de estilo desse autor são as que ocorreram em pelo menos 75% dos textos (ou seja, em pelo menos 15, dos 20 textos estudados).

### **3.2 ANÁLISE DO NÍVEL NARRATIVO**

Para a Semiótica, todo texto apresenta uma narratividade, um nível narrativo, subjacente. Assim, a narratividade, neste programa de pesquisa, não está ligada, conforme se costuma pensar, apenas aos textos que estão organizados narrativamente em sua superfície e relatam ações ou eventos. Ao contrário: a todo texto subjaz uma relação entre um “sujeito” e um “objeto”, que pode ser conjunta ou disjuntiva, e algumas vezes ocorrem transformações nessa relação; são essas relações e transformações que constituem a narratividade de um texto. Quando um sujeito está em disjunção com um objeto, é essa falta que faz “andar” o texto. Tais transformações são organizadas em etapas que a semiótica francesa convencionou chamar de “esquema narrativo canônico” (BARROS, 2003). Esse esquema prevê que todo texto pressupõe três etapas:

- *manipulação*: fase em que um destinador tenta convencer um destinatário a fazer ou crer em algo;
- *ação*: fase em que o sujeito age para mudar o seu estado conjunto;
- *julgamento*: fase em que um destinador julgador avalia as ações e

comportamentos dos sujeitos.

A palavra-chave aqui é “pressupõe”. Isso significa dizer que nem todas as etapas previstas precisam estar presentes em um texto, mas elas estão logicamente pressupostas. Se existe uma sanção, é porque alguma ação existiu e, se esta existe, é porque houve uma manipulação. Existem textos que estarão concentrados na etapa da ação; outros, como alguns textos publicitários, por exemplo, podem apresentar só a manipulação. Além disso, essas etapas podem ser organizadas de modos diferentes, não sendo necessária a sua aparição em ordem lógica (FIORIN, 2005, p. 34).

Os textos podem ainda se desenrolar a partir de um “acontecimento” ou de uma “rotina” (ZILBERBERG, 2007). O acontecimento é algo que se dá de repente, um fato raro que ocorre sem que o sujeito o espere e que o tire de sua rotina. Ele pode ser responsável por iniciar uma ação por parte do sujeito (fazer missivo) ou por arrebatá-lo, privando-o do agir (fazer remissivo). Quando um texto é da ordem da rotina, nada de novo acontece, há espaço para trivialidades. Acontecimento e Rotina estão ainda relacionados a outro nível de significação: o nível tensivo (ZILBERBERG, 2006a; 2006b; 2006c), descrito mais adiante.

- *Programas narrativos/Esquema narrativo canônico*

Como já visto, a narratividade subjacente a todos os textos, conforme o modelo proposto pela Semiótica, envolve transformações no estado de um sujeito com relação a um objeto. Quando o que se vê é apenas um estado juntivo entre um sujeito e um objeto (ou seja, apenas se identifica a situação do sujeito com relação ao objeto – conjunção ou disjunção), diz-se que existe um enunciado de estado; quando ocorre uma transformação nesse estado, existe um enunciado de fazer. Quando um enunciado de fazer muda um enunciado de estado, ou seja, algo ocorre e isso altera a situação do sujeito com relação a um objeto, existe um programa narrativo<sup>26</sup>. Um texto pode apresentar apenas um programa narrativo ou vários, e esses programas podem ser classificados segundo vários critérios. Barros (2003, pp. 21-26) explica tais critérios e apresenta diversos exemplos. Os programas narrativos ocorrem dentro das diferentes etapas do “esquema narrativo canônico” (manipulação, ação, julgamento). O Quadro 3, a seguir, mostra esse esquema:

---

<sup>26</sup> Segundo Barros (2003, p. 20), um programa narrativo é “um enunciado de fazer que rege um enunciado de estado”.

ESQUEMA NARRATIVO CANÔNICO									
MANIPULAÇÃO				COMPETÊNCIA/AÇÃO				SANÇÃO	
percurso do destinador manipulador				percurso do sujeito				percurso do destinador julgador	
PN de doação de competência semântica		PN de doação de competência modal		PN de competência		PN de performance		PN de interpretação	PN de retribuição
EN do fazer S1 ≠	EN de estado S2	EN do fazer S1 ≠	EN de estado S2	EN do fazer S1 ≠	EN de estado S2	EN do fazer S1 =	EN de estado S2	EN do fazer S1 ≠	EN de estado S2

**Quadro 3:** Esquema narrativo canônico.

PN = percurso narrativo; EN = enunciado.

Fonte: Barros (2003).

Pode-se ver o esquema narrativo canônico como uma constante, porém o modo como suas etapas se articulam é variável. Conforme Barros (2003, p. 43), o esquema narrativo (...) oferece muitas oportunidades de variações e combinações que dão caráter único e específico às narrativas-ocorrências”. Ora, se as combinações, como diz Barros, dão caráter único às narrativas-ocorrências, supõem-se que recorrências de tais combinações ao longo de diferentes textos de um mesmo autor podem ser uma marca de autoria, de um estilo.

A maior parte dos textos do autor em questão não apresenta o esquema narrativo canônico completo, mas concentra-se apenas na fase da sanção, em que o sujeito (destinador-julgador) avalia a ação de outros sujeitos. A fase da manipulação também ocorre em muitos textos, mas escolheu-se aqui como característica mais marcante a sanção, que é mais frequente. Verifica-se que dentre 20 textos do mesmo grupo, 18 (90%) apresentam a fase da sanção, num percurso de interpretação. Em muitos desses textos, observa-se também que a sanção é a fase principal em torno da qual se organiza o texto. No Quadro 4, a seguir, encontram-se trechos dos textos que corroboram essa análise. Observa-se que o enunciador julga, avalia ações e comportamentos dos sujeitos.

Texto	Comentário	Trechos retirados do texto
1	Ao procurar saber se vão ou não poder usar um certo cupom para jogar boliche, o enunciador ocupa uma posição de julgador, já que avalia o comportamento e o estado dos amigos que se tornaram pais recentemente.	- Afinal, quero saber: agora que você já é mãe e o Jonas já é pai, já devem saber a resposta. - Eu acho que ogroupon dos pais está perdido.... Nao largarão da pequena!!! hihhi
2	A fase da sanção pode ser percebida imediatamente, já que o enunciador faz uma avaliação negativa do gosto do enunciatário.	- não gostei nada nada do filme... -Xinguei-te mentalmente algumas vezes -aqueles que vc emprestou eu gostei!!!” -nao gostei do chapeuzinho vermelho sexy

3	Avalia-se a conduta de outro sujeito. Pede-se ao enunciário que cobre de uma determinada empresa os pagamentos em atraso referentes a um acordo judicial e, ao fazê-lo, avalia o comportamento da empresa negativamente.	<i>(o que pode ser pressuposto por meio de “já” e “por duas vezes”): - “por duas vezes, em novembro e dezembro, a Reclamada já havia depositado o valor do acordo com atraso”</i>
4	O sujeito avalia negativamente a ação de deputados que votaram a favor de emenda constitucional com a qual não concorda. Há também uma avaliação negativa do país.	<i>- Massssssssss “se liga” na nossa querida CAMARA DOS DEPUTADOS (uso irônico do adjetivo 'querida'). - ABSURDOOOOOOOO (sobre a emenda) - Como nosso Brazilão é ^%@^\$%&amp;*@^(*#</i>
5	O enunciador avalia os comportamentos daqueles que se divorciam.	<i>- Alguns entram ali apenas para conversar e contar os problemas - Outros ali ingressam com muita paz e harmonia</i>
8	Sujeito avalia situação do meio ambiente, bem como a ação da "humanidade", despreocupada com este.	<i>- Este tema foi escolhido devido aos reflexos da conduta humana no meio ambiente, que tem gerado a longo prazo, grandes estragos ambientais, comprometendo assim, a sadia qualidade de vida (...) a humanidade parece não se preocupar imediatamente com este problema, haja vista que seus reflexos surgem ao longo de anos.</i>
9	Sujeito avalia negativamente a opinião das pessoas que gostam do frio.	<i>- (pergunta) “precisa passar frio ou congelar?????????” - (afirmação jocosa) “Este frio só para a menopausa mesmo!”</i>
10	A contar que alguém jogara filhotinhos de gato dentro de sua casa, o sujeito avalia negativamente quem cometeu essa ação. Avalia também a ação da gatinha-mãe, que enfrentara o cão bravo.	<i>- O que o gênio não sabia é que eu tinha um cachorro extremamente violento, que estava solto no quintal - a teimosa e valente mini-gatinha resolveu enfrenta-lo, ficando em duas patinhas</i>
11	Ao saber que a criança está doente, avalia-se que isto pode ter se dado porque ela quer alguma coisa.	<i>- Sú... as vezes as crianças ficam doentes quando querem alguma coisa... Será que não significa que a Sonia quer um dalmata ou um gatinho malhadinho?? : o )</i>
12	Assim como nos outros textos, aparece um momento em que o enunciador faz uma avaliação, ocupando a posição de destinatador-jogador, na fase da sanção.	<i>- Lembrei da viagem pro xxxxxxx pro casamento dos gaúchos Fulana e Beltrano (que foi lindo!)</i>
13	Sujeito avalia negativamente a OAB-SP	<i>- A colônia que a OABSP tem, é lá no fim do mundo virando a esquerda, né? - Caso você não saiba, se você tentar cancelar sua OABSP em Jaaneiro, vc deve pagar o ano INTEIRINHO, mesmo sem utilizar. - (em um momento, descreve várias qualidades da OAB-MG, o que desqualifica a OAB-SP) - acho que a de SP está precisando de um bom chacoalhão, hein?</i>
15	Sujeito avalia seu próprio conhecimento.	<i>- meu inglês está enferrujadaaaaço.</i>
16	Há implicitamente uma avaliação negativa dos organizadores de um concurso, que mudaram as regras "no meio do jogo".	<i>- Esse é o video de dança da minha amiga. Eles estão competindo no dancetube. Eram primeiro lugar, mas o concurso foi prorrogado e faltando 5 dias estão quase perdendo o primeiro lugar (que já era deles se</i>



		<i>tivesse encerrado na data prometida)</i>
17	O pilar do texto é a surpresa pela descoberta do “verdadeiro” café, em contraposição ao café do supermercado. Há claramente uma avaliação dos tipos de café.	- Houve uma época em que eu achava que aquilo que vendiam no mercado em embalagem de café, era café!!! Hoje sei: <u>aquilo é palha!</u> - Café é só o da Tia Ivone, em que ela mesma planta, torra e mói!
18	Uma avaliação é evidente, uma vez que, mesmo ao se colocar à disposição para receber tais gatos como doação, o enunciador não deixa de avaliar negativamente as pessoas que compram gatos de raça.	- Acho sem-vergonhice comprar gato de raça, com tantos SRD's por aí
19	Há uma avaliação negativa dos serviços que encerram suas atividades às 16 ou 17h.	-Não dá pra resolver tudo em 3 dias, ainda que fique até tarde (pois a maioria dos rolos são em cartórios, bancos, etc, que fecham as 16 ou 17h! E ainda, pdem 5 dias uteis para emitir certidoes)
20	Neste texto, mesmo ao tentar convencer o enunciatário a fazer um procedimento, o sujeito admite que este é trabalhoso.	-Vai dar um pequeno trabalhinho
21	Sobre um bolo que um amigo fez, num primeiro momento, o sujeito o avalia positivamente; depois, julga que ele deveria ter sido feito de outra forma.	- Está lindo - Mas vai uma dica: pega um pedaço de linha, apoia a mão na metade do bolo e corta o bolinho em dois, porém, mantendo o formato redondo. Corta na altura e nao na metade dele!!! Rssss

**Quadro 4:** Nível narrativo – esquema narrativo: fase da sanção; exemplos.

Foi dito anteriormente que os textos podem ainda se desenrolar a partir de um “acontecimento” ou de uma “rotina” (ZILBERBERG, 2007), sendo o primeiro algo que ocorre repentinamente que tire o sujeito de sua rotina. Ele pode ser responsável por iniciar uma ação por parte do sujeito (fazer missivo) ou por arrebatá-lo, privando-o do agir (fazer remissivo). Quando um texto é da ordem da rotina, nada de novo acontece, há espaço para trivialidades. No caso dos textos analisados, sua maioria é da ordem do acontecimento (e não da rotina) o qual desencadeia um fazer missivo por parte do sujeito. Dos 20 textos, 15 (75%) apresentaram essa característica. Veja o Quadro 5 para exemplos extraídos dos textos.

<b>Texto</b>	<b>Comentário Geral/Exemplos</b>
1	Os amigos tornaram-se pais e isso suscita uma dúvida quanto a poderem ou não ir jogar boliche num futuro próximo. Essa dúvida surge a partir do momento em que os amigos tornam-se pais: “ <i>Afinal, quero saber: agora que você já é mãe e o Jonas já é pai, já devem saber a resposta. É possível deixar a pequena por algumas horinhas para jogar boliche daqui uns meses ou isso é muito ilusório?!? Rsssss</i> ”. Este fato raro, é um “acontecimento” nos termos de Zilberberg (2007) e vem impedir o sujeito de continuar o seu programa (jogar boliche com os amigos). Pode-se dizer que o acontecimento causou uma ruptura, uma quebra na continuidade.
2	O fato de o enunciatário ter gostado do filme pega o enunciador de surpresa (de novo, acontecimento), expressada por “ <i>Nossa</i> ” em “ <i>Nossa Jane...</i> ”.

3	O enunciador deixa claro sua surpresa em constatar que a Reclamada mais uma vez atrasou o pagamento, pois ela já tinha sido avisada: “ <i>Contudo, mesmo após sua orientação, a empresa novamente deixa de depositar na data acordada</i> ”. Assim, mais uma vez o texto é da ordem do acontecimento, já que instaura uma ruptura na rotina do sujeito, que tem que tomar a ação de escrever o e-mail de cobrança.
6	O texto se inicia com a constatação, por parte do enunciador, de que faltam dados num processo de inventário em que trabalha. Essa falta é o que cria uma ruptura, já que o sujeito não pode continuar o seu trabalho. Para suprimir a falta, ele escreve ao enunciatário. “ <i>Boa tarde, Mário! Tudo bom? Precisamos de alguns dados faltantes para a conclusão do inventário</i> ”.
8	A ruptura que faz com que o sujeito componha o texto é a percepção de que (grifo nosso) “ <i>Atualmente estes reflexos da degradação ambiental tem se demonstrado de maneira mais clarividente</i> ”.
9	O sujeito não esperava o frio intenso e é afetado por ele, ficando surpreso, como pode ser visto em: “ <i>menina do céu (...). Eu ganhei umas tulipas... Li que vc tinha que guardar o bulbo no congelador para simular a situação real dela na natureza. Se eu soubesse que ia esfriar desse jeito, tinha deixado na varanda mesmo!!!</i> ”
11	O que motiva a escritura do enunciado é a doença da filha da amiga (acontecimento).
13	O acontecimento é a notícia da construção de uma colônia de férias da OAB: “ <i>Outro dia eu vi uma reportagem que a OAB ganhou EM 2006, um terreno pra construir uma colônia de férias em Campos de Jordão</i> ”.
15	Neste texto vê-se um sujeito que é tirado de seu dia a dia ao assumir uma nova ocupação: a de professora de inglês. Trata-se de uma surpresa, já que não se esperava a proposta para o cargo, como se vê em “ <i>Obrigada!!! Especialmente a Tá, que surgiu com esta possibilidade</i> ” (grifo nosso).
16	Algo totalmente inesperado aconteceu: a prorrogação de um concurso (fato não muito comum) o que tirou a liderança da amiga.
17	O pilar do texto é a surpresa pela descoberta do “verdadeiro” café: “ <i>Houve uma época em que eu achava que aquilo que vendiam no mercado em embalagem de café, era café!!! Hoje sei: aquilo é palha!</i> ”
21	Bolo feito por um amigo, de quem não ela não esperava tal ação. O texto ressalta explicitamente que se está surpreso com o que foi feito pelo sujeito “Julio”: “ <i>Meu.... me amarrota que eu to passada!</i> ”

**Quadro 5:** Acontecimento/Fazer missivo.

### 3.3 ANÁLISE DE ASPECTO, MODALIDADES E PAIXÕES

A Semiótica vê as paixões como efeitos de sentido resultantes das combinações, dos arranjos entre as modalidades (querer, dever, poder, saber) que incidem sobre o “ser”. Além disso, elas se distinguem de acordo com tipo de objeto da junção, com a temporalidade e com a aspectualidade.

Na esfera passional, pode-se dizer que, na maioria dos textos do autor analisado, vê-se um sujeito do inconformismo que é ou aquele que não-quer-criar e/ou aquele que não se resigna. Inconformado com o acontecimento, o sujeito não se resigna, não “deixa para lá”. Ao contrário: ele escreve sobre aquilo e demonstra por vezes a sua indignação. O inconformismo e/ou a indignação são quase sempre a motivação para a composição do texto. É isso que faz o enunciador escrever. Dos 20 textos que constituem o grupo, 18 (90%) apresentam a paixão do inconformismo. Veja, no Quadro 6, trechos extraídos do texto que corroboram o inconformismo,

algumas vezes elevado a uma indignação:

Texto	Comentário	Trechos retirados do texto
1	Sujeito inconformado, que não espera, não aceita passivamente os acontecimentos, o que o motiva a escrever o enunciado para tentar suprimir a situação de ignorância em que vive. Ele não aceita não saber o que vai acontecer com o cupom. Quer saber o quanto antes para poder tomar uma decisão. Evidente pela <b>cobrança</b> que faz aos amigos.	- <i>Afinal, quero saber: agora que você já é mãe e o Jonas já é pai, já devem saber a resposta. É possível deixar a pequena por algumas horinhas para jogar boliche daqui uns meses ou isso é muito ilusório?!?</i> Rsssss - <i>Precisamos saber para gastarmos o groupon: o boliche rola ou não rola até antes de 28/08?!?</i>
2	O sujeito não se conforma com o fato da amiga ter gostado de um certo filme (não quer crer no gosto dela). Alia-se ao inconformismo certa indignação pelo gosto do outro (evidenciada por “ <i>xinguei-te mentalmente</i> ”).	- <i>Nossa Jane... não gostei nada nada do filme... Xinguei-te mentalmente algumas vezes, tá?</i>
3	O sujeito não se conforma com a persistência nos atraso. Escreve ao enunciatário cobrando.	- <i>Contudo, mesmo após a sua orientação, a empresa novamente deixa de depositar na data acordada.</i> - <i>por duas vezes, em novembro e dezembro, a Reclamada já havia depositado o valor do acordo com atraso.</i>
4	Inconformismo/Indignação com relação à tentativa de aprovação de uma Emenda Constitucional	- <i>conjunção “Mas” (ainda reforçada pelo fato de ter sido escrita com duplicação de letras); uso irônico do adjetivo “querida”; uso de caixa alta para escrever todas as palavras da frase em que conta o que está para acontecer (caixa alta é associado, na Internet, a uma ênfase ou a um modo de indicar que se está “gritando”):</i> <i>“Massssssssss “se liga” na nossa querida CAMARA DOS DEPUTADOS:”</i> - <i>uso de “absurdo”; explicitação da concordância com os partidos que votaram contra a emenda: “ABSURDOOOOOOO (...). Únicos que votaram contra: PPS, PV e PSOL!!!!!!! (Ganharam meu voto!!)”</i>
8	O texto é a Introdução de um Trabalho de Conclusão de Curso, cujo tema é a proteção ao direito dos animais. O sujeito inconformado e que avalia negativamente as ações humanas sobre o meio ambiente fica evidente pelos exemplos da coluna ao lado.	- <i>Este tema foi escolhido devido aos reflexos da conduta humana no meio ambiente, que tem gerado a longo prazo, grandes estragos ambientais, comprometendo assim, a sadia qualidade de vida (...)</i> - <i>Estes reflexos da degradação ambiental tem se demonstrado de maneira mais clarividente. Contudo, a humanidade parece não se preocupar imediatamente com este problema, haja vista que seus reflexos surgem ao longo de anos.</i>
9	O enunciador não quer acreditar que existem pessoas gostam do frio.	- <i>Eu nunca entendi essa ideia de ‘curtir o frio’ - precisa passar frio ou congelar?????????</i>

10	Ao contar que alguém jogara filhotinhos de gato dentro de sua casa, diz “ <i>O que o gênio não sabia é que eu tinha um cachorro extremamente violento, que estava solto no quintal</i> ”, ou seja, não se conforma com a ingenuidade do praticante da ação. Percebe-se também que justamente a ação avaliada positivamente pelo sujeito é a ação de um sujeito da não conformidade: a gatinha não se resigna diante do perigo: ao contrário, parte para a ação.	- <i>O que o gênio não sabia é que eu tinha um cachorro extremamente violento, que estava solto no quintal</i> - <i>a teimosa e valente mini-gatinha resolveu enfrenta-lo, ficando em duas patinhas</i>
12	Mesmo após ter sido picado por insetos e estar com as pernas inchadas, o enunciador, que não se conforma, não se rende, vai à aula de dança.	- <i>Meu pé voltou parecendo um pãozinho depois da picada... Mas deu pra dançar numa perna só!! : o</i>
13	Sujeito não se conforma com certas “regras” da OAB-SP.	- <i>A colônia que a OABSP tem, é lá no fim do mundo virando a esquerda, né?</i> - <i>Mas nunca mais vi nada a respeito.</i> - <i>Depois de conhecer outras OAB’s, acho que a de SP está precisando de um bom chacoalhão, hein?</i>
15	A paixão do inconformismo aparece novamente, não aliada a uma indignação mas como sentido do sujeito inconformado como aquele que não se resigna. Mesmo com o inglês “ <i>enferrujadaaaaço</i> ”, vai “ <i>se jogar</i> ” na nova atividade”.	- <i>Deu um friozinho na barriga de assistir as aulas com as crianças que não param quietas (rss) mas vamos se jogar!!!!hihihi</i> - <i>E estudar!!! Pq meu inglês está enferrujadaaaaço..!</i>
16	O inconformismo é claro com as regras do concurso que mudaram.	- <i>estão quase perdendo o primeiro lugar (que já era deles se tivesse encerrado na <u>data prometida</u>)</i>
17	O sujeito inconformado aparece novamente, com um tom de indignação, por exemplo, em “ <i>Fui enganada a vida toda!!!</i> ”. E ele não se conforma, não aceitando o café que lhe foi apresentado a vida toda e se rende ao “verdadeiro” café: “ <i>Agora sim eu posso dizer que gosto de café – mas do café-café! Não do pseudo-café!</i> ”	- <i>Fui enganada a vida toda!!!</i>
18	Considerando que uma das acepções de inconformismo é “tendência ou atitude de não se acatar passivamente o modo de agir e de pensar da maioria do grupo em que se vive” ( <i>Dicionário Houaiss</i> ), verifica-se que o sujeito não se conforma com o fato de algumas pessoas pagarem para ter gatos.	- <i>Acho sem-vergonhice comprar gato de raça, com tantos SRD’s por aí</i>
21	Sujeito não se conforma, não quer crer, que certo amigo possa ter feito um bolo tão bom.	- <i>Meu.... me amarrota que eu to passada! O Julio que fez???? Nao é nada do que imaginei!!! –Faz outro pra mim????</i>

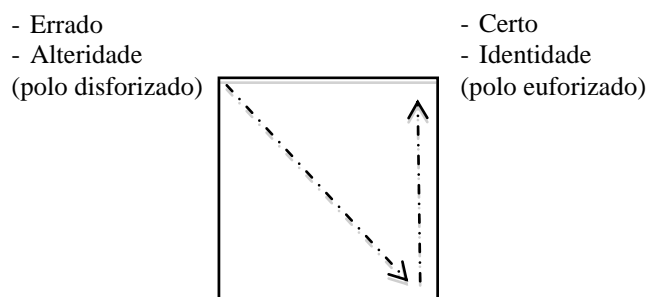
**Quadro 6:** Paixão do inconformismo.

### 3.4 ANÁLISE DO NÍVEL FUNDAMENTAL

Para a Semiótica Discursiva, os objetos que os sujeitos buscam no nível narrativo estão investidos de valores correspondentes a termos de uma categoria semântica que se instaura num nível mais profundo: o fundamental. Por exemplo,

um preso que foge tem na fuga um objeto investido do valor de “Liberdade” oposto à “Opressão”, sendo essa oposição uma categoria semântica básica que vai permear todo o texto. Essa categoria se articula em termos de negação ou afirmação de seus polos. As relações entre os dois polos são dadas por um modelo lógico que representa as relações de contradição, contrariedade e complementaridade entre os termos: o quadrado semiótico. Outros exemplos de categorias semânticas básicas frequentemente analisadas quando se utiliza a Semiótica como referencial teórico são “vida e morte”, “natureza e cultura”, “identidade e alteridade”. Cada um dos termos de uma categoria semântica básica é um valor positivo ou negativo, dentro de cada texto.

Nos textos em questão, existe a recorrência da categoria semântica fundamental do “errado” *versus* “certo”. O Quadro 7<sup>27</sup>, a seguir, apresenta algumas passagens encontradas no *corpus* que exemplificam que, ao julgar negativamente certas ações “dos outros”, o enunciador cria a oposição mencionada, valorizando o “certo” identificando-o como o que ele faz e desvalorizando o errado, que é sempre a atitude ou o comportamento alheio. Assim, cria também mais uma oposição semântica: “alteridade” *versus* “identidade”, com a valorização do “eu” e a desvalorização do “outro”. O quadrado semiótico, na Figura 3 homologa essas relações.



**Figura 3:** Quadrado semiótico: “certo *vs.* errado” / “identidade *vs.* alteridade”.

<b>Texto</b>	<b>Exemplos</b>
2	- <i>Nossa Jane... não gostei nada nada do filme... Xinguei-te mentalmente algumas vezes, tá? (hauhauhuaha).</i>
3	- <i>por duas vezes, em novembro e dezembro, a Reclamada já havia depositado o valor do acordo com atraso.</i> - <i>Solicitei ao Dr. que orientasse a Empresa para que não mais houvesse atraso, havendo de minha parte certa tolerância quanto aos atrasos.</i>
4	- <i>Como nosso Brasília é ^%@^\$%&amp;*@^(*#, até hoje existem "donos de cartório" por</i>

<sup>27</sup> Os grifos encontrados são nossos e foram feitos quando, em trechos muito grandes, há partes menores que se relacionam às categorias mencionadas.

	<i>"herança de família", sem concurso público. ELES QUEREM APROVAR UMA EMENDA CONSTITUCIONAL (UMA "LEI", PARA OS LEIGOS) PROPONDO QUE, QUEM ESTÁ LÁ SEM CONCURSO, TENHA O DIREITO DE FICAR!!!! ABSURDOOOOOOO.</i>
5	<i>- Discute-se cada armário comprado - e neste momento você percebe a mesquinha em que vivem alguns casais, que costumam guardar notas fiscais de anos atrás, já pensando em uma futura separação. Buscam partilhar aquilo que não é seu de direito - apenas pelo objetivo de uma "falsa vingança". Os filhos viram simples moeda de troca. Enquanto a mãe, que geralmente detem a guarda, busca o direito de pensão alimentícia, o pai, apenas para tirar-lhe a paz, ingressa com pedido de visitas, mesmo sabendo que irá retirar a criança e deixá-la com a avó, sem manter qualquer contato. E a situação perdura. Perdura de forma indeterminada. Conselho e consulta gratuita da Dra.: não estão felizes? Façam sim o divórcio. Todos tem direito à felicidade.</i>
8	<i>- Neste mesmo capítulo, veremos as manifestações populares que configurariam o delito em estudo, como a vivisseção, rodeios, circos e a farra- do-boi. - Este tema foi escolhido devido aos reflexos da conduta humana no meio ambiente, que tem gerado a longo prazo, grandes estragos ambientais, - Atualmente estes reflexos da degradação ambiental tem se demonstrado de maneira mais clarividente. Contudo, a humanidade parece não se preocupar imediatamente com este problema.</i>
9	<i>- Eu nunca entendi esta ideia de "curtir o frio". Ficar debaixo do cobertor ou sair na rua parecendo um boneco de neve é curtir o frio?? Eu hein!!! (...). Este frio só para a menopausa mesmo! Se você tiver esta alegria por "curtir o frio" (e eu sei que vc é destas), vc vai amar aqui!!!</i>
10	<i>- Quando eu tinha 08 anos, alguém jogou dentro da minha casa três filhotinhos de gatos... todos cheios de pulga! O que o gênio não sabia é que eu tinha um cachorro extremamente violento, que estava solto no quintal. - Adote um animal: a felicidade está por aí...! : o ) (o enunciador conta a história de como conseguiu sua gata de estimação, valorizando como é bom ter um animalzinho). - Se precisar, estou na loucura, mas conte mesmo comigo! (Texto 11)</i>
13	<i>- A OABMG dá uma "remuneração" para advogada em licença-maternidade, isenção da anuidade no ano do nascimento do bebê, máquina de suco e café expresso nas salinhas e, ao cancelar sua anuidade, vc paga proporcional ao mês de cancelamento. Caso você não saiba, se você tentar cancelar sua OABSP em Janeiro, vc deve pagar o ano INTEIRINHO, mesmo sem utilizar. - Enfim. Logo, logo teremos eleições OABSP. Depois de conhecer outras OAB's, acho que a de SP está precisando de um bom chacoalhão, hein?</i>
16	<i>- Eram primeiro lugar, mas o concurso foi prorrogado e faltando 5 dias estão quase perdendo o primeiro lugar (que já era deles se tivesse encerrado na data prometida).</i>
17	<i>- Fui enganada a vida toda!!! Agora sim eu posso dizer que gosto de café - mas do café-café! Não do pseudo-café!</i>
18	<i>- Acho sem-vergonhice comprar gato de raça, com tantos SRD's por aí. Mas, por outro lado, sempre quis um desses. Então, vai que alguém vai pro exterior e precise deixar o gato com alguém que goste destes bichinhos... estou por aí!! =^..^=</i>
19	<i>- (pois a maioria dos rolos são em cartórios, bancos, etc, que fecham as 16 ou 17h! E ainda, pdem 5 dias uteis para emitir certidoes).</i>
20	<i>- Ah vai.... eu sempre ajudo quando pedem! Vamos ajuda-lo!!</i>
21	<i>- Mas vai uma dica: pega um pedaço de linha, apoia a mão na metade do bolo e corta o bolinho em dois, porém, mantendo o formato redondo. Corta na altura e nao na metade dele!!! Rssss</i>

**Quadro 7:** Exemplos de passagens do corpus que remetem às oposições “certo vs. errado” / “identidade vs. alteridade”.

### 3.5 ANÁLISE DO NÍVEL TENSIVO

Conforme visto anteriormente, no nível tensivo, articulam-se as categorias

mais abstratas dos textos, tais como continuidade *versus* descontinuidade ou relaxamento *versus* tensão. O nível tensivo pode ser considerado o mais abstrato na imanência de um texto. Os temas e figuras do nível discursivo, as transformações do nível narrativo, as paixões, o aspecto pressupõem a articulação do sentido em categorias mais gerais, tais como “relaxamento” e “tensão”, ou “continuidade” e “descontinuidade”.

Outra noção importante a considerar no que diz respeito à tensividade é a de extensidade *versus* intensidade. Zilberberg (2006a, pp. 169-170) propõe que a tensividade seja a união dessas grandezas; a primeira une a temporalidade e a espacialidade, e a última, o andamento e a tonicidade. A intensidade diz respeito ao sensível, aos estados de alma, e a extensidade, ao inteligível, aos estados de coisas.

Vê-se que, em termos de intensidade, a maioria dos textos opera em uma tonicidade forte, já que existe um sujeito inconformado (ver seção 3.3), que está sempre avaliando (na maior parte das vezes, negativamente) as ações dos outros (conforme visto na seção 3.2) e cuja motivação para compor os enunciados é um acontecimento (conforme seção 3.2, ou seja, algo que o tira de seu dia a dia, compõe textos calcados na tensão. Se um acontecimento é o que está na base dos enunciados, esses textos são também da ordem da descontinuidade. No entanto, em muitos dos textos, verifica-se que o enunciador procura de algum modo “amenizar” tal tensão. Dos 20 textos que constituem este grupo, 17 apresentam essa tensão junto com uma tentativa de amenização, conforme pode ser conferido com alguns exemplos extraídos do *corpus* e que se encontram no Quadro 8.

<b>Texto</b>	<b>Comentário Geral</b>	<b>Atenuação da Tensão (ou Tensão Amenizada)</b>
1	O fato de os amigos terem se tornado pais é uma quebra na continuidade. o sujeito percebe o descompasso que há entre o seu andamento (acelerado) e o do mundo (dos outros, dos enunciatários). Daí a cobrança (“ <i>rola ou não rola (...)?!?</i> ”; “ <i>É possível deixar a pequena (...) ou isso é muito ilusório?!?</i> ”). Essa cobrança eleva a tonicidade do enunciado.	A tensão é amenizada pelas indicações constantes de riso (“ <i>rsssss</i> ”, “ <i>hihihi</i> ”).
2	Assim como no texto 1, uma tensão se instaura, já que pode ser criada uma polêmica a respeito da qualidade do filme, principalmente por se tratar de uma avaliação do gosto de amigos.	Aqui também o sujeito vai atenuar essa tensão utilizando alguns marcadores como risos (“ <i>hauhauhaha</i> ”, “ <i>hihi</i> ”), o “ <i>tá?</i> ” e um diminutivo (“ <i>Bjinhos</i> ”).
3	O ocorrido tira o sujeito de seu estado de relaxamento, há uma quebra na continuidade. Ele deve fazer uma cobrança, o que instaura uma tensão no texto.	Tentativa de amenizar a tensão, porém, pelo fato de o texto ser mais formal, utilizam-se estratégias diferentes das vistas até o momento, como o uso de “ <i>peço a gentileza</i> ”.

4	Existe uma tonicidade forte, indicada também por elementos do plano da expressão tais como o uso de caixa alta e repetição de exclamações.	Tensão amenizada pela inclusão de expressões jocosas e uso de ironia, como "trem da alegria", "se liga", "querida câmara"
5	O divórcio, tema deste texto, é um acontecimento, que tira os envolvidos de estado relaxamento e os coloca em um estado tenso, sobre o qual o enunciador discorre: "O desgaste emocional que gera um divórcio é imensurável".	A tensão devido ao tema delicado é amenizada também neste texto, por meio do tom mais informal do fim do texto, principalmente pelo início do último parágrafo: "Conselho e consulta gratuita da Dra" e pela expressão "sacudam a poeira", bastante coloquial.
8	A ruptura que faz com que o sujeito componha o texto é a percepção de que (grifo nosso) "Atualmente estes reflexos da degradação ambiental tem se demonstrado de maneira mais clarividente". Este estado do meio ambiente tira o sujeito de seu relaxamento.	Tensão se ameniza já que o sujeito estudará o meio ambiente, tão importante: "Sendo assim, iremos abranger neste trabalho este bem jurídico tão importante, de uso comum do povo: o meio ambiente".
10	O fato de ter colocado os gatinhos pulguentos sobre a cama é inesperado e estabelece uma tensão, uma descontinuidade.	Tensão suavizada pelo riso ("hihi"). Além disso, o cachorro é citado, logo no início do texto, criando-se uma tensão que é em seguida amenizada por "Um anjo nos fez perceber isso (...)".
11	O Sujeito é tirado de seu estado de relaxamento pela doença da filha da amiga; descontinuidade.	O fato motivador do enunciado é algo um tanto grave – a doença da criança –, mas há também aqui a atenuação da tensão, por meio pergunta em tom de brincadeira envolvendo as figuras dos animais de estimação e uso do diminutivo: "Será que não significa que a Sonia quer um dalmata ou um gatinho malhadinho??".
12	Ocorreu um acontecimento (picada de insetos) que instaura descontinuidade.	Novamente, vê-se a tensão do enunciado e a gravidade da situação sendo amenizadas pelo tom de brincadeira de algumas expressões, como: "inchaço das perninhas!!!", "Meu pé voltou parecendo um pãozinho depois da picada... Mas deu pra dançar numa perna só!! : o"
13	Ocorre descontinuidade e tensão ao receber resposta da OAB-SP e constatar diferenças com relação à OAB-MG.	O texto começa tenso com um tom formal "Colegas advogados paulistas" e termina relaxado com "Beijinhos".
16	O pedido para que outros assistam ao vídeo acaba por ser quase uma ordem, ao se usar o imperativo, ainda reforçado pela exclamação "assistam!".	Novamente se vê uma atenuação da tensão com o uso do diminutivo "rapidinho".
19	A tensão é instaurada pela constatação de que não se conseguirá resolver todas as atividades.	Tensão amenizada pelo riso e tom jocoso ("Pelo visto vou resolver parcialmente tudo e resolver integralmente nada! rssss A advocacia ainda me pira!").
20	Sujeito faz um pedido. A tensão/urgência pode ser depreendida logo no início: "Gente!!!!".	Tensão amenizada ao se usar o diminutivo duas vezes para explicar que a tarefa que o enunciário terá que executar não é trabalhosa: "Vai dar um pequeno trabalhinho".
21	O texto ressalta explicitamente que se está surpreso (descontinuidade) com o que foi feito pelo sujeito "Julio": "Meu.... me amarrota que eu to passada!".	Julga que o bolo deveria ter sido feito de outra forma. Avaliação negativa (crítica) que instaura tensão, que, por sua vez, é amenizada pelos risos ("rsssss").

**Quadro 8:** Tonicidade forte/descontinuidade; atenuação da tensão.



### 3.6 NÍVEL DISCURSIVO

O nível discursivo é onde estão os temas e as figuras que recobrem os valores dos níveis fundamentais e tensivo e as transformações da estrutura narrativa. Neste nível, “todas as formas abstratas do nível narrativo são revestidas de termos que lhe dão concretude” (FIORIN, 2005, p. 41). Dentre os níveis do plano do conteúdo, o nível discursivo é, portanto, o mais concreto e também o mais variável. Por exemplo, o estado de um Sujeito em disjunção com um Objeto, próprio do nível narrativo, pode ser revestido de muitas formas no nível discursivo; pode aparecer como alguém que quer uma casa e não a tem, ou um animal que quer um dono e não o tem, ou uma princesa que quer se casar, etc. Neste nível se depreendem ainda as projeções da enunciação no enunciado, através das categorias de pessoa, tempo e espaço.

A Semiótica observa que o nível discursivo é o mais variável dos níveis no que diz respeito à semântica, pois as categorias semionarrativas e tensivas, mais abstratas, podem ser revestidas por numerosas figuras diferentes que lhe dão concretude (esse é o nível mais próximo da manifestação). Um mesmo enunciador pode utilizar as mesmas estruturas semionarrativas ou tensivas para tratar de assuntos tão diferentes quanto um problema de trabalho ou animais que ele viu no parque, por exemplo. Assim, este nível não foi levado em consideração neste momento por esta pesquisa. Posteriormente, pretende-se verificar se a sintaxe do nível discursivo (instauração das categorias de pessoa, tempo e espaço) também pode funcionar como marcador de estilo.

### CONCLUSÃO E COMPARAÇÃO COM O TEXTO DE AUTORIA QUESTIONADA

Em resumo, as características do plano do conteúdo que se mostraram relevantes na análise devido a sua recorrência, como marcadores do estilo do autor são aquelas que figuram na Tabela 1, a seguir:

	<b>Total de textos por característica</b>	<b>%</b>
Sanção - percurso de interpretação	18	90%
Acontecimento	15	75%
Inconformismo	18	90%
“Certo” x “Errado”	15	75%
Tonicidade forte/Descontinuidade ; Atenuação da Tensão	17	85%

**Tabela 1:** Marcadores de estilo do autor analisado (plano do conteúdo)

O fato de um mesmo conjunto de características do plano do conteúdo ser encontrado ao longo da maior parte dos textos do mesmo autor, mesmo que esses textos tenham sido bastante diferentes quanto à sua natureza, corrobora a hipótese inicial de que há menos variação intrafalante nos níveis mais profundos dos textos.

Foi analisado também o texto que, neste exercício, funciona como texto questionado, para se verificar se essas características estão presentes. O Quadro 9 a seguir, apresenta o resultado da análise.

<b>Característica</b>	<b>Comentário</b>	<b>Alguns exemplos extraídos do texto</b>
Sanção	Avalia o comportamento da amiga, sua reação e as atitudes com relação à celebração de seu casamento.	- "nunca vi um casamento tão detalhado como o seu"; - "não é comum cada detalhe que vc pensou!!" - "Desde pequena gosta de organizar cada detalhe, como fazíamos nos trabalhos de maquetes e musicas. Para alguns isso é 'nóia'. Para nós sempre foi 'o trabalho bem feito'. - "muito pelo contrario, vc está bem calma (...) será o casamento da princesa, pq vcs pensaram em tudo para ser o casamento do sonho" - "Bom.... acho que vc já ficou brava"
Acontecimento	O sujeito constata o planejamento de um casamento detalhado como nunca visto; - Há uma má reação da amiga frente à brincadeira realizada.	- "Nao entenda mal a brincadeira." - "Mas acredite se quiser, não é comum cada detalhe que vc pensou!!" - "Bom.... acho que vc já ficou brava"
Inconformismo	Não se conforma com todos os preparativos para o casamento.	- "não é comum cada detalhe que vc pensou!!" - " Mas sendo sincera, não é comum! Eu, pelo menos, nunca vi." - "São trocentos mil detalhes que vc pensa!! Eu acho que vc não dorme!!!!" - "resolvi falar para nao ficar mal explicado"
Tonicidade forte / Descontinuidade; Atenuação da Tensão	A tensão é instaurada pelas críticas que faz à amiga. Indicações de riso ao longo do texto tentam amenizar a situação, mais a despedida carinhosa (ver ao lado).	- "rsss" - "Beijos" - "amo vcs"

**Quadro 9:** Características do texto questionado.

A repetição da mesma combinação de traços indica que é muito provável que o autor do texto questionado seja o mesmo dos textos conhecidos (de fato, neste exemplo, sabe-se que é).

O estilo se dá pela recorrência de um conjunto, de uma cosseleção de características (no plano da expressão ou do conteúdo); nunca por uma ou outra característica tomada isoladamente. Assim, é claro que uma ou outra dessas

características podem aparecer em textos de outros autores, mas *é muito improvável que exatamente a mesma combinação de todos esses elementos esteja presente.*

Dando continuidade a esta pesquisa, trabalhos futuros deverão incluir:

- exercícios de análise com outros autores e, dentre eles, fazer um “teste cego”, ou seja, testar a autoria de textos em que realmente não se saiba quem é o autor;

- modelos de análise quantitativa que confirmam mais suporte às conclusões da análise qualitativa.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. ALMEIDA, D. *Semiótica da poesia: estudo de poemas de Paulo Henriques Britto*. 2009. 125 f. Dissertação (Mestrado em Linguística) – Universidade de São Paulo, São Paulo, 2009.
2. BAAYEN, H. et al. An experiment in authorship attribution. *Journées internationales d'analyse statistique des donnees textuelles*, n.6, 2002.
3. BARROS, D. L. P. *Teoria do discurso: fundamentos semióticos*. São Paulo: Humanitas, 2001.
4. \_\_\_\_\_. *Teoria semiótica do texto*. São Paulo: Ática, 2003.
5. BELL, A. Language style as audience design. *Language in society*, n.13, p.145-204, 1984.
6. BLOCH, B. A set of postulates for phonemic analysis. *Language*, n. 24, p.3-46, 1948.
7. BUTTERS, R. Sociolinguistic variation and the law. In: BAYLEY, Robert; LUCAS, Ceil. *Sociolinguistic variation: theories, methods, and applications*. New York: Cambridge University Press, 2007.
8. \_\_\_\_\_. Forensic Linguistics. *Journal of English linguistics*, v.2, n. 39, 2011.
9. CHASKI, C. Who's at the keyboard? Authorship attribution in digital evidence investigations. *International journal of digital evidence*, v.4, n.1, 2005.
10. \_\_\_\_\_. Recent validation results for the syntactic analysis method for author identification. In: INTERNATIONAL CONFERENCE ON LANGUAGE AND LAW, Cardiff, Wales, 2004.
11. \_\_\_\_\_. Empirical evaluations of language-based author identification techniques. *Forensic linguistics*. Birmingham: University of Birmingham, v.8, n.1, 2001.
12. \_\_\_\_\_. Who wrote it? Steps toward a science of authorship identification. *National Institute of Justice journal*, sep. 1997.

13. CORNEY, M. *Analysing e-mail text authorship for forensic purposes*. 2003. 165 f. Tese (Master of Information Technology) – Queensland University of Technology, Brisbane, 2003.
14. COULTHARD, M; JOHNSON, Alison. *An introduction to forensic linguistics: language in evidence*. London/New York: Routledge, 2007.
15. COULTHARD, M. ... *and then...* *Language description and author attribution*. 2006. Disponível em <<http://www1.aston.ac.uk/lss/staff/coulthardm>>. Acesso em 28 jul. 2010.
16. \_\_\_\_\_. Linguistas como peritos/as. *Linguagem em (Dis)curso*, Tubarão, v. 4, n.esp: 159-176, 2004a.
17. \_\_\_\_\_. *Author identification, idiolect and linguistic uniqueness*. 2004b. Disponível em <<http://www.business-english.ch/downloads/Malcolm%20Coulthard/AppLing.art.final.pdf>>. Acesso em 28 jul. 2010.
18. \_\_\_\_\_. Early famous cases of disputed authorship. INTERNATIONAL SUMMER SCHOOL IN FORENSIC LINGUISTIC ANALYSIS, 12, Barcelona. *Informação oral/Slides de aula*. 2012.
19. \_\_\_\_\_. On the use of corpora in the analysis of forensic texts. *International Journal of Speech Language and the Law*, v. 1, n.1, 1994.
20. COULTHARD, M.; HAGEMEYER, C. Perigo, cuidado, atenção: a comunicação linguística de risco em advertências de produtos. *Cadernos de Linguagem e Sociedade*, v. 2, n. 14, 2013.
21. COUPLAND, N. *Style: language variation and identity*. Cambridge: Cambridge University Press, 2007.
22. DISCINI, N. *O estilo nos textos*. São Paulo: Contexto, 2009.
23. ECKERT, P; RICKFORD, J. R. *Style and sociolinguistic variation*. Cambridge: Cambridge University Press, 2001.
24. FBI (2008). *FBI 100: the Unabomber*. Disponível em <[http://www.fbi.gov/news/stories/2008/april/unabomber\\_042408](http://www.fbi.gov/news/stories/2008/april/unabomber_042408)>. Acesso em 19/02/2103.
25. FIORIN, J. L. *Em busca do sentido: estudos discursivos*. São Paulo: Contexto, 2008.
26. \_\_\_\_\_. *Elementos de análise do discurso*. São Paulo: Contexto, 2005.
27. GIBBONS, J. Towards a framework for communication evidence. *The international journal of speech, language and the Law*, Equinox, v.18, n.2, 2011.

28. \_\_\_\_\_. Approaching questions in forensic authorship analysis. In: GIBBONS, John; TURELL, M. Teresa (eds.). *Dimensions of forensic linguistics*. Amsterdam; Philadelphia, PA: John Benjamins, 2008.
29. \_\_\_\_\_. Quantifying evidence in forensic authorship analysis. *The international journal of speech, language and the Law*, Equinox, v.14, n.1, 2007.
30. \_\_\_\_\_. *Authorship attribution in a forensic context*. 2004. 410 f. Tese (PhD) – University of Birmingham, Birmingham, 2004.
31. \_\_\_\_\_. TXT 4N6: method, consistency, and distinctiveness in the analysis of SMS text messages. *Journal of law and policy*, n.2, v. XXI, p.467-494, 2013.
32. GREIMAS, A. J.; COURTÉS, J. *Sémiotique: dictionnaire raisonné de la théorie du langage*. Paris: Hachette, 1979.
33. \_\_\_\_\_.; \_\_\_\_\_. *Dicionário de Semiótica*. São Paulo: Cultrix, 2008.
34. GREIMAS, A. J.; FONTANILLE, J. *Semiótica das paixões: dos estados de coisas aos estados de alma*. São Paulo: Ática, 1993.
35. HARKOT-DE-LA-TAILLE, E. *Ethos e autoria*. In: SIMPÓSIO INTERNACIONAL SOBRE ANÁLISE DO DISCURSO: EMOÇÕES, ETHOS E ARGUMENTAÇÃO, III, 2008, Belo Horizonte. *Anais*. Belo Horizonte: UFMG, 2008. v.1.
36. HJELMSLEV, L. *Prolegômenos a uma teoria da linguagem*. São Paulo: Perspectiva, 2003 [1943].
37. HOLMES, D. The evolution of Stylometry in Humanities scholarship. *Literary and Linguistic Computing*, vol. 13, n. 3, 1998.
38. \_\_\_\_\_. A stylometric analysis of Mormon Scripture and related texts. *Journal of the Royal Statistical Society*, n. 155, p. 91-120, 1992.
39. HONORÉ, A. Some simple measures of richness of vocabulary. *ALLC Bulletin*, v.2, n.7, 1979.
40. IRVINE, J. 'Style' as distinctiveness: the culture and ideology of linguistic differentiation. In: ECKERT, Penelope; RICKFORD, John R (eds.). *Style and sociolinguistic variation*. Cambridge: Cambridge University Press, 2001.
41. JUOLA, Patrick. Questioned electronic documents: empirical studies in authorship attribution. In: OLIVIER; SHENOI (eds.). *Research advances in digital forensics II*. Heidelberg: Springer, 2006.
42. \_\_\_\_\_. Idiolect in forensic authorship attribution. *Folia linguistica anglica*, Lodz: University of Lodz Press, 4, 2002.

43. \_\_\_\_\_. Defining the scope of forensic linguistics/Introduction to forensic authorship analysis. INTERNATIONAL SUMMER SCHOOL IN FORENSIC LINGUISTIC ANALYSIS, 12, Barcelona. *Informação oral/Slides de aula*. 2012.
44. \_\_\_\_\_. *Forensic Linguistics and the status of linguistic evidence in the legal setting*. 2000. Tese (PhD) – University of Lodz, Lodz, 2000.
45. LABOV, W. *The social stratification of English in New York City*. São Paulo: Cambridge University Press, 2006 [1966].
46. \_\_\_\_\_. The isolation of contextual styles. In: *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press, 1972.
47. LILLIS, T. *The sociolinguistics of writing*. Edinburgh: Edinburgh University, 2013.
48. LOVE, H. *Attributing authorship: an introduction*. Cambridge: Cambridge University, 2002.
49. McMENANIN, G. R. Forensic stylistics: theory and practice of forensic stylistics. In: COULTHARD, M.; JOHNSON, A. (eds.). *The Routledge handbook of forensic linguistics*. London; New York: Routledge, 2010.
50. \_\_\_\_\_. *Forensic Linguistics: advances in forensic stylistics*. Boca Raton, Florida, USA: CRC Press, 2002.
51. OLSSON, J. *The language detective: forensic linguistics first certificate course*. The Forensic Linguistics Institute, 2008. Livro digital.
52. \_\_\_\_\_. *Forensic Linguistics: an introduction to language, crime, and the Law*. London: Continuum, 2004.
53. RUDMAN, J. The state of authorship attribution studies: some problems and solutions. *Computers and the Humanities*, Kluwer Academic, 31, p. 351-365, 1998.
54. SAPIR, E. Speech as a personality trait. *American journal of Sociology*, 32: 892-905, 1927.
55. SCHILLING-ESTES, N. Investigating stylistic variation. In: CHAMBERS, J.K.; TRUDGILL, P.; SCHILLING-ESTES, N. (orgs). *The handbook of language variation and change*. Oxford; Malden, MA: Blackwell, 2001.
56. SILVA, R. Listas de frequência de palavras como marcadores de estilo no reconhecimento de autoria. In: LINGUATECA: 10 ANOS, 2008, Aveiro. *Actas*. Disponível em < <http://www.linguateca.pt/LivroL10/>>. Acesso em 23 ago. 2010.
57. SILVA, R. et al. Comparing sentence-level features for authorship analysis in Portuguese. In: 9th INTERNATIONAL CONFERENCE PROPOR, 2010, Porto Alegre. *Proceedings*. Springer, 2010.

58. SOLAN, L.; TIERSMA, P. *Speaking of crime: the language of criminal justice*. Kindle edition. Chicago; London: The University of Chicago, 2005.
59. TATIT, L. *Análise semiótica através das letras*. São Paulo: Ateliê Editorial, 2001.
60. TIERSMA, P. *What is forensic linguistics?* s/d. Disponível em <<http://www.languageandlaw.org/FORENSIC.HTM>>. Acesso em 25 ago. 2013.
61. TOMBLIN, S. Investigating formulaic language as a marker of authorship. In: THE INTERNATIONAL ASSOCIATION OF FORENSIC LINGUISTS' TENTH BIENNIAL CONFERENCE, 2011, Birmingham. *Proceedings*. Birmingham: Centre for Forensic Linguistics, 2012.
62. TURELL, M. T. Idiolectal similitude (or distance) in authorship analysis. INTERNATIONAL SUMMER SCHOOL IN FORENSIC LINGUISTIC ANALYSIS, 12, Barcelona. *Informação oral/Slides de aula*. 2012.
63. \_\_\_\_\_. The use of textual, grammatical and sociolinguistic evidence in forensic text comparison. *The international journal of speech, language and the Law*, Equinox, n.17, v.2, p.211-250, 2010.
64. WINTER, E.; WOOLLS, D. *Identifying authorship in a co-written novel*. Internal report for the University of Birmingham, 1996.
65. WOOLLS, D; COULTHARD, M. Tools for the trade. *Forensic Linguistics*. Birmingham: University of Birmingham, n.5, v.1, 1998.
66. ZILBERBERG, C. Louvando o acontecimento. *Revista Galáxia*, Tradução de Maria Lucia Vissotto Paiva Diniz. São Paulo, n. 13, p. 13-28, jun. 2007.
67. \_\_\_\_\_. Síntese da gramática tensiva. *Revista Significação*, São Paulo: Annablume, n.25, 2006a.
68. \_\_\_\_\_. *Razão e poética do sentido*. Tradução de Ivã Carlos Lopes; Luiz Tatit; Waldir Beividas. São Paulo: Edusp, 2006b.
69. \_\_\_\_\_. *Eléments de grammaire tensive*. Limoges: Pulim, 2006c.

**ABSTRACT:** Asking “Who wrote this text?” has been frequent in some textual analysis when it comes to literary, religious or historical texts. More recently, the same question emerged in forensic contexts, in order to help solve crimes and judicial processes. This paper aims at giving an overview of the area of authorship analysis for forensic purposes, as well as pointing out possible problems and how they could be solved

**Keywords:** Authorship attribution; Discourse Semiotics; Style; Forensic Linguistics.

Recebido no dia 20 de junho de 2014.

Aceito para publicação no dia 20 de agosto de 2014.