



Os porquês de Wanda Tinasky não ser Thomas Pynchon – um estudo de caso (II)

The reasons why Wanda Tinasky not be Thomas Pynchon: a case study (II)

Saulo Brandão^(a)

^a Universidade Federal Rural de Pernambuco, Pernambuco, Brasil – saulo.brandao@ufrpe.br

Resumo: Neste trabalho, voltamos a enfrentar o problema sobre a autoria das cartas de Wanda Tinasky. Cartas essas que foram dadas como de autoria do ficcionista estadunidense Thomas R. Pynchon; tese que foi desmentida por Donald Foster (2000) baseada em evidências históricas. Em 2018, escrevemos um primeiro ensaio fazendo a mesma contestação de Foster, mas baseado em dados referentes às escolhas lexicais de cada escritor; ou seja: uma refutação por meios linguísticos. Neste trabalho, enfrentamos, de novo, o problema para ratificar o resultado de Foster e do nosso trabalho anterior; mas, desta vez, optamos por analisar valores como comprimento de períodos, frequência de sinais ortográficos e, especialmente, a riqueza lexical de cada corpus. Consideramos que o resultado confirmou nossa hipótese e, por fim, acreditamos, restar provado linguisticamente que Pynchon não escreveu as cartas de Tinasky.

Palavras-chave: Thomas Pynchon. Wanda Tinasky. Neolo. Cartas. Estilometria.

Abstract: In this work we return to face the problem about the authorship of Wanda Tinasky's letters. These were given as authored by the American fictionist Thomas R. Pynchon; hypothesis that was denied by Donald Foster (2000) based on historical evidence. In 2018, we wrote a first essay making the same contest as Foster but based on data pertaining to each writer's lexical choices. That is: a refutation by linguistic means. In this work, we face the problem again to ratify Foster's result and that of our previous work, but this time we chose to analyze values such as length of periods, frequency of orthographic signs and, especially, the lexical richness of each corpus. We consider that the result confirmed our hypothesis and, finally, we believe, it rests linguistically proven that Pynchon did not write the letters of Wanda Tinasky.

Keywords: Thomas Pynchon. Wanda Tinasky. Neolo. Cartas. Stylometry.

Introdução

Temos como objetivo neste texto negar, por meios linguísticos, a participação de Thomas R. Pynchon (1937) no episódio conhecido como *As cartas de Wanda Tinasky*, uma série de cartas publicadas no *Anderson Valley Advertiser (AVA)* no período de 1983 a 1988. Isto porque, em 1990, Bruce Anderson (editor da AVA) aventou a possibilidade de Wanda ser um pseudônimo utilizado pelo festejado romancista estadunidense. A partir deste ponto, os estudiosos e fãs de Pynchon começaram uma jornada para ligar ou negar o ficcionista à Wanda Tinasky e suas cartas.

Para chegar aos nossos objetivos, neste texto, utilizamos o software **Neolo** (Brandão; Crogwey, 2014) para analisar as cartas de Tinasky e comparar os resultados obtidos com: a) os resultados observados ao examinar peças dos textos não-ficcionais de Pynchon; b) os resultados de seus textos ficcionais e c) as características de estilo de outras cartas publicadas pela AVA em resposta às cartas de Wanda para, a partir desses resultados, emitirmos um parecer conclusivo favorável ou não às declarações de Bruce Anderson.

Observando os resultados, nosso leitor notará peculiaridades aparentes nos vários corpora sob investigação e, especialmente, nos textos da prosa não-ficcionais de Pynchon quando comparados com *As cartas*. Essa diferença é a que mais importa para esta pesquisa. Este texto também reanimará a ideia de que cada um dos textos tem uma dicção linguística que funciona como uma impressão digital. Em outras palavras: as escritas produzidas por uma pessoa têm características próprias que não são encontradas em textos da lavra de outra pessoa. Quando escrevemos, tendemos a usar um conjunto particular de palavras, ou sequências de palavras, ou estruturas de sintaxe próprias, nossas. Essa ideia povoou as investigações de estudiosos desde a fundação da Biblioteca de Alexandria

em aproximadamente 300 a.C. (Love, 2002, p. 23), sendo que a disputa, daquela época, que mais nos toca nesta empreitada foi a distinção feita entre a *Iliada* e a *Odisseia*, como sendo obras de Homero, das outras odisseias épicas envolvendo a guerra de Tróia. Desde então, pesquisadores têm desenvolvido estratégias e ferramentas para lidar com esse fenômeno, principalmente com o objetivo de descobrir os autores de textos anônimos ou apócrifos. Essas primeiras averiguações só foram possíveis após a leitura deixar de ser privilégio de escribas e com a grande Biblioteca de Alexandria, que iniciou a tradição de colecionar, em um único local, escritos diversos e de origens diferenciadas. Essas práticas chegaram ao medievo pelas mãos do clero católico interessado em definição de autorias de livros e textos bíblicos, especialmente os dos evangelistas. E, naturalmente, as análises ganharam o espaço secular para analisar todos os tipos de textos que despertassem dúvidas sobre suas autorias.

Os fundamentos da estilometria moderna foram estabelecidos pelo filósofo polonês Wincenty Lutosławski em *Principes de stylométrie* (1890). Lutosławski usou esse método para desenvolver, com sucesso, uma cronologia dos *Diálogos de Platão*.

Antes, em 1851, Mendenhall tentou caracterizar o estilo de diferentes autores através da distribuição de frequência de palavras de vários comprimentos. Em sua pesquisa, ele verificou a possível relevância dessa técnica para a questão da autoria em Shakespeare e, vários anos depois, a suspeita foi enfrentada para verificar se Francis Bacon era o verdadeiro autor das obras geralmente atribuídas a Shakespeare. A diferença entre Mendenhall e Lutosławski é que este estabeleceu uma metodologia para identificação de autoria a partir do estilo e aquele apenas utilizava uma ferramenta da depois fundada estilometria.

Na modernidade, os estudos estilométricos foram densamente escrutinados e centenas de textos foram analisados às minúcias. Contudo, no início dos anos 2000, esses estudos e metodologias foram perdendo popularidade e hoje são mais frequentes nas mãos de investigadores da área da Estatística, muito mais preocupados em desenvolver metodologias e fórmulas estatísticas mais robustas e complexas do que as anteriores (Olsson, 2008; Pennebaker; King, 1999). Os estatísticos se valem de grandes corpora literários estruturados, ou não, depois de devidamente limpos e balizados para seus exercícios. Entretanto, seu objetivo passa ao largo da historiografia e crítica literária.

Pynchon em suma

Que Pynchon é uma figura significativa entre os escritores de romance contemporâneos nos Estados Unidos da América, não há dúvidas, dado ao número de grandes prêmios literários que coleciona e pela vendagem de seus livros. É de conhecimento dos que labutam com literatura anglófona o seu hábito recluso, sendo o autor lembrado pelos críticos por sua afiliação a um grupo de escritores que filtram seus leitores, apenas aceitando aqueles de destacada erudição¹. Por conta disso, foi categorizado como escritor enciclopédico junto com Joyce, Cervantes, Dante (Mendelson, 1976, p. 161). E, do lado do leitor, diz-se preciso ter uma enciclopédia por perto para fazer uma apreensão razoável dos textos de Pynchon, ou ter uma fusão de horizonte de expectativas muito afinada com o do escritor para o sucesso da leitura e interpretação de seus textos.

Por causa de seu estilo, algumas palavras são muito frequentes em textos críticos sobre as obras do autor e seu léxico, sintaxe e discurso. Aqui estão algumas dessas palavras/conceitos: caótico, cômico, invectivo,

¹ Essa estratégia é revelada e analisada no *Livro Pós-escrito a O nome da rosa* de Umberto Eco. Sendo o teórico e ficcionista italiano um dos que, como Pynchon, adotam esse filtro.

esquizofrênico, paródico, kitsch, fetiche, inventivo, fragmentado, paranoico, ciência/científico (tudo, desde os gregos até Ilya Prigogine e Higgs), entropia, Demônio de Maxwell, Gato de Schrödinger e muitas menções à cultura popular como Godzilla, King Kong, Simpsons etc. Lidando com um léxico muito estranho e uma gramática muito sofisticada, Pynchon desenvolveu seus nove romances entre 1963 e 2013, e ainda continua escrevendo. Com essa percepção pelos críticos, é desafiador encontrar escritores que compartilhem com ele seus dicionários, estratégias e sintaxe. Alguns críticos veem um arranjo próximo do estilo de Pynchon nos livros de David Foster Wallace (1962-2008), especialmente *Infinity jest* (1996). Porém, até mesmo nos textos de Wallace, embora semelhantes aos de Pynchon, um leitor atento pode identificar as diferenças com clareza.

Com essas já mencionadas características, não é fácil arremedar textos de Pynchon. No entanto, para alguém tentar fazê-lo, seria preciso um alto grau de sofisticação para falar sobre tantos tópicos tão profundamente e com pertinência. Dessa forma, uma pessoa disposta a imitar o estilo de Pynchon não só tem que lidar com o estilo, o vocabulário, a visão cínica da sociedade e assim por diante, mas ser uma pessoa muito erudita. Seria um trabalho insano, por nada, uma vez que uma pessoa com essa sagacidade e habilidade poderia se tornar um escritor original.

Escrevemos esses parágrafos apenas como uma breve introdução sobre Pynchon para qualquer um de nossos potenciais leitores que não saiba muito sobre o autor entenderem o propósito deste texto, que é: linguisticamente, separar Thomas Pynchon do tumultuado caso Wanda Tinasky.

Quem foi Wanda Tinasky?

De 1983 a 1988, o jornal AVA recebeu uma série de cartas assinadas por alguém que se nomeava Wanda Tinasky – *The Fort Bragg bag lady*. Tais cartas foram escritas por uma pessoa erudita que era capaz de criticar escritores e poetas e outros artistas, políticos e possuía uma verve bem-humorada e satírica. Era, pois, o tipo de escrita sarcástica, cômica e sofisticada tão festejada nos textos de Pynchon.

Em 1990, Bruce Anderson descobriu que Pynchon viveu no norte da Califórnia quando o AVA recebeu e publicou as cartas de Tinasky. O ficcionista estava escrevendo seu quarto romance, *Vineland* (1990), cujo enredo acontece na mesma região. A partir dessa descoberta, Anderson lançou ao público a possível conexão entre Tinasky e Pynchon.

Após essa suspeita levantada, os editores do AVA, Anderson, Fred Gardner e T. R. Factor, começaram a investigar a relação entre Pynchon e a ‘*bag lady*’. Gardner entrou em contato com a agente de Pynchon, Melaine Jackson, que negou que o escritor fosse o autor das cartas e negou o uso do nome de Pynchon de qualquer forma relacionada com as cartas. Neste momento, Gardner desistiu do projeto Tynasky. Foi Factor quem manteve o projeto em execução e publicou, em 1996, um livro com as cartas de Wanda Tinasky. Nele, Factor coletou as cartas de Tinasky, também muitas outras cartas de pessoas que concordavam e discordavam da posição da ‘*bag lady*’. Steven Moore escreveu o prefácio do livro e emprestou toda a sua autoridade como editor, crítico literário e autor respeitado para o reconhecimento das cartas de Tinasky como sendo uma produção de Pynchon.

Aí vem Donald Foster

Donald Foster é um estudioso shakespeariano que, em sua tese doutoral, se propunha a identificação do autor de uma elegia assinada por um W. S., que ele provou ser abreviação de William Shakespeare. Depois disso,

Foster trabalhou em muitos casos como um investigador linguístico ou literário forense. Em seu fólio, havia determinações autorais do romance *Primary colors* (1996), que ele indicou corretamente como sendo Joe Klein; a disputa Theodore Kaczynski/Unabomber; o caso de assassinato de Jonbenét Ramsey; o ataque de antraz; o processo de Monika Lewinsky/Paula Jones. Contudo, Foster nem sempre se provava certo e houve casos que ainda não foram resolvidos. Mas, sobre a elegia de W.S., reconheceu, em 2006, que estava errado: Shakespeare não havia escrito o poema.

Em 1996, Klein admitiu o resultado de Foster reconhecendo ser ele o autor de *Primary colors* e, no mesmo ano, Foster foi convidado a trabalhar na definição da pessoa que escreveu as cartas de Wanda Tinasky. Ele levou dois anos para propor uma resposta, que estava longe de ser uma descoberta linguística. Ele revela em seu livro *Author unknown: On the trail of anonymous* (2000) que T. D. Factor, que estava executando um projeto para vincular as cartas de Tinasky a Pynchon, convidou-o a examinar o material que ele havia coletado do AVA; não somente as cartas de Tinasky, mas também outras que ele pensou poderem ser provenientes da mesma lavra.

Foster nunca usou publicamente métodos linguísticos para negar a participação de Pynchon na escrita das cartas de Tinasky. Seus argumentos não eram muito conclusivos ou óbvios: 1) Tinasky era muito ácida em suas críticas aos artistas, e Pynchon é muito generoso sempre que escreve sobre outros artistas; 2) Tinasky mencionava a política partidária e histórias literárias de uma época em que Pynchon era uma criança; e, 3) Tinasky citou muitos poetas da Geração Beat que viviam no Noroeste dos EUA. Esta informação, para Foster, colocava a *'bag lady'* na costa oeste dos EUA e como sendo uma pessoa muito mais velha do que Pynchon.

Vimos esses achados com parcimônia, uma vez que, para um escritor inventivo como Pynchon, não seria difícil incorporar fatos históricos e questões éticas de outros períodos em seus escritos – sendo isso tudo que ele faz em suas ficções. O próprio Foster declarou que tais dados eram circunstanciais, não conclusivos. Ele havia se posicionado contra Pynchon ser a *'bag lady'*, mas precisava de indícios mais assertivos para firmar sua posição final.

A pista certa veio em uma das cartas de Tinasky em é mencionado o nome de Gary Snyder (1930). Nessa carta, ela alegou ter escrito um poema para saudar o nascimento do filho de Snyder, há cerca de 18 anos àquela época, e mencionou alguns versos do poema. Snyder, nos anos 90, era professor na Universidade da Califórnia – Davis, em Sacramento. Foster o enviou um e-mail perguntando sobre o poema e mencionou os versos que Wanda havia revelado em sua carta. Snyder respondeu dizendo que se lembrava do poema e que Tom Hawkins escrevera a peça.

No passo seguinte, Foster ligou para um número de telefone listado no guia telefônico para um chamado Tom Hawkins no condado de Mendocino – cidade do AVA. A pessoa disse que ela não era quem Foster estava procurando, mas que lembrava das notícias sobre a morte de um Tom Hawkins, mais velho. Ou seja, a pessoa listada no guia não era o Tom Hawkins em questão. Aquele relatou que este morreu em um acidente de carro depois de matar sua esposa e incendiar sua casa.

Foster descobriu o endereço da casa do Tom Hawkins mais velho e conversou com os novos donos. O casal, que morava lá, foi muito cooperativo. Eles descreveram e até enviaram a Foster algum material encontrado em um galpão perto da casa queimada. Lá, eles encontraram a máquina de escrever Underwood, que Hawkins costumava usar para escrever as cartas e correspondências endereçadas ao AVA. Os tipos da

máquina apresentavam falhas coincidentes com as observadas nos originais das cartas de Tinasky.

Como se pode ver, não há provas linguísticas que tenham levado Donald Foster a descartar Pynchon como autor das cartas de Tinasky. De fato, começamos a revelar tais provas linguísticas em um texto anterior publicado, também, na revista *Texto Digital*² e que concluiremos neste escrito.

Etapas da pesquisa

Primeiro, vamos rerepresentar o software que usamos para analisar os textos. Seu nome é *Neolo*, e nós o desenvolvemos especialmente para fazer alguns cálculos que outros algoritmos não fazem, mas também alguns que são comuns - mas era importante tê-los em um mesmo software por conta da lógica de programação que às vezes gera discrepância entre os resultados de um algoritmo e outro. Embora sua função principal seja extrair o neologismo de um documento, não usamos essa função neste estudo³. Além dessa função primária, *Neolo* pode desempenhar outras 14 funções. E, neste trabalho, utilizamos: a) seis das funções que determinam a riqueza do texto⁴ (três dessas funções são sensíveis ao tamanho do corpus e três não são); b) a função que estabelece a frequência dos sinais de pontuação nos textos; e, c) aquela que determina o tamanho médio das frases (em número de palavras por frase).

Depois que digitalizamos todo o texto e os transformamos no modo .TXT (*Neolo* só reconhece texto simples, sem estruturas internas), verificamos os tamanhos dos textos que pretendíamos comparar. As cartas contam com

² [Vista do Os porquês de Wanda Tinasky não ser Thomas Pynchon – um estudo de caso \(I\) \(ufsc.br\)](#)

³ Esse dado foi utilizado no primeiro texto que publicamos na *Texto Digital*, citado acima.

⁴ Riqueza lexical e variedade lexical podem ser usados de forma intercambiáveis em estilometria, sendo o segundo termo mais explícito de seu significado, mas o primeiro tem a preferência dos pesquisadores da área.

um pouco mais de 28.000 palavras e, por isso, decidimos estabilizar todos os textos com aproximadamente o mesmo tamanho das cartas. Escolhemos as primeiras 28.000 palavras de *Vineland*. A razão pela qual escolhemos esse romance foi porque era ele que Pynchon estava escrevendo no mesmo tempo que o AVA estava publicando as cartas. O outro corpus em uso foi um formado por textos não-ficcionais escritos por Pynchon – Pynchon leva sua reclusão tão a sério que foi com dificuldade que conseguimos juntar esse pequeno corpus de textos não-ficcionais escritos por ele. Fizemos o mesmo com outras cartas publicadas pela AVA relacionadas com as de Tynasky. Desta forma, trabalhamos com quatro corpora diferentes. Renomeamos os arquivos assim:

| | |
|----------------------------------|---------------|
| Vineland | Vineland |
| Textos não-ficcionais de Pynchon | Pynonfic |
| Cartas de Wanda | Tinasky |
| Outras cartas na AVA | Other_Letters |

É essencial informar aqui que todos os textos são maiores que 28.000 palavras e menores do que 28.300. Deixamo-los um pouco maiores porque, dependendo do editor de texto ,a contagem é um pouco diferente, mas não o suficiente para mudar as conclusões que encontramos.

Usamos as seguintes ferramentas da **Neolo**:

1. Número de períodos no texto (reflete a frequência na utilização sinais de pontuação como: ponto final, reticências, sinal de interrogação e de exclamação)
2. Comprimento médio da frase (em número de palavras)

3. Número de hápax legonema⁵
4. TTR (proporção Tipo - Token)
5. HTR (razão Hápax - Token)
6. HTyR (razão Tipo - Hápax)
7. MSTTR (Média segmental da relação Tipo - Token)
8. MTLT (Medida da Diversidade Léxica Textual)
9. HD-D

Como consideramos o nome de várias funções listadas acima como autoexplicativas (1, 2, 3, 4, 5, 6, 10 e 11), vamos agora explicar, de forma breve, as ferramentas listadas em 7, 8 e 9, por não serem óbvias⁶.

Johson propôs, em 1944, a de número 7. Ela é de fácil compreensão: o programa divide o texto em segmentos de 100 palavras e, em seguida, faz com que o TTR (Proporção Tipo-Token) de cada segmento e, finalmente, o MSTTR é obtido através do cálculo da média aritmética dos TTR de todos os segmentos. O de número 8 (McCarthy, 2005) funciona de forma semelhante ao de número 7, mas os segmentos são formados seguindo uma regra lógica e não um número fixo de palavras. Acontece assim: quando o TTR de um segmento atinge 0,72, o programa quebra uma seção e inicia uma nova. MTLT é o resultado de 'L/n', onde 'L' é o comprimento do texto no número de palavras e 'n' é o número de segmentos obtidos. HD-D é de cálculo diferente: o programa seleciona uma palavra aleatoriamente e, em seguida, calcula qual é a possibilidade desta palavra estar entre 42 (padrão fixo) outras palavras escolhidas aleatoriamente. Dado ao grande número de segmentos formados, a média não é de todos os resultados, mas

⁵ Palavra ou expressão que ocorre apenas uma única vez no texto (conceito bom para a área de Textometria).

⁶ Explicações detalhadas sobre esses mecanismos podem ser obtidos no artigo *Lexical statistics and typological Structures: A measure of lexical richness* de de Torruella e Capsada em: <https://www.sciencedirect.com/science/article/pii/S1877042813041888>.

dos escolhidos seguindo uma regra estatística chamada distribuição de probabilidades hipergeométricas. Essas três ferramentas foram concebidas para que se pudesse comparar a riqueza lexical de textos de tamanhos diferentes – embora não seja o caso dos corpora aqui utilizados. Elas entraram na pesquisa por serem mais precisas quando lidando com corpora pequenos. Essas três ferramentas foram explicadas acima na ordem de sofisticação. Portanto, a HD-D é a mais sofisticada e leva uma vantagem muito grande sobre as demais, porque quebra a lógica da escolha dos seguimentos de forma sequencial. Existe um verdadeiro embaralhamento estatístico das palavras para depois avaliar a possibilidade de elas aparecerem num grupo de 42 outras palavras também escolhidas aleatoriamente.

Decidimos usar a riqueza lexical do texto como principal índice para indicar autoria ou não dos corpora, porque ela varia muito de escritor para escritor. Utilizando essa média de comparação, juntamente com as outras ferramentas que listamos anteriormente, encontramos algumas diferenças significativas entre os autores e até mesmo entre os dois corpora de Pynchon (fictício e não-fictício). No entanto, a descoberta mais crítica baseia-se na ideia de que cada um de nós tem uma impressão digital escrita, isso significa que, inconscientemente, usamos certas palavras de uma maneira diferente das outras pessoas. Halteren *et al.* (2005) usaram o termo *stylome* (estilome, doravante) para dar um nome específico em vez de emprestar o termo 'impressão digital' de outras áreas. Tivemos que investigar vários métodos para encontrar um estilome.

Nossa preocupação também foi utilizar software livre (gratuito) para que outros pesquisadores linguistas ou estudiosos literários pudessem replicar nossa pesquisa e adotar a metodologia para investigar corpora de seus interesses. Foi por essa razão que desenvolvemos o software **Neolo**

reunindo exatamente as ferramentas⁷ que utilizamos nessa investigação sobre a *As cartas de Wanda Tinasky*. Tomamos todo cuidado para não envolver a pesquisa com detalhes estatísticos que fogem, no mais das vezes, da compreensão dos humanistas. A metodologia é simples e o software é livre e está disponível em <https://github.com/jcrowley/neolo>. Mas o pesquisador pode sempre fazer quase os mesmos passos da pesquisa utilizando outros softwares livres como *Lexico5* ou *Antconc*, com a exceção de as ferramentas não estarem disponíveis em um único software. E, até onde sabemos, o HD-D não é encontrado em programas gratuitos. O método que aplicamos foi uma versão simplificada da proposta por Peng e Hengartner (2001).

Um parêntese terminológico

Embora creiamos que este artigo será majoritariamente lido por pessoas que tenham iniciação em estudos estilométricos e, para esses, a explicação seja desnecessária, temos que prever que, eventualmente leitores alheios a esse viés de pesquisa se interessem pelo tema ou pelos sujeitos escritores dos corpora aqui utilizados, portanto, precisamos fazer este aparte.

Os dois termos que acreditamos mereça atenção são *Type* (T) e *Token* (TR), que podem ser traduzidos como Tipo e Palavra (em uma tradução de contexto, a rigor a *token* se traduz por signo ou sinal), respectivamente.

Esses vocábulos têm que ser bem entendidos para a compreensão de algumas formulações, como, por exemplo, a ferramenta de número 4, acima: TTR, que se traduz matematicamente como T/TR e isso quer dizer que o número de Tipos dividido pelo número de Palavras, onde os Tipos são as palavras que formam o dicionário do texto e Palavras são todas as

⁷ As outras 3 ferramentas de *Neolo* foram utilizadas no primeiro texto que publicamos sobre o assunto, já anteriormente referenciado.

palavras que, em tecitura, formam o texto. Ou seja: as Palavras são os Tipos repetidos diversas vezes no texto.

Uma pergunta pode nascer de um leitor atento: mas qual é a diferença de um Tipo para um hápax legomena? Sim, parece ser uma dúvida legítima! Observem que, quando definimos os hápax na nota 4, dissemos que estes só aparecem uma única vez no texto, enquanto os Tipos se repetem diversas vezes no texto.

Vejamos esses exemplos:

O vocábulo *VOW* é uma hápax legomena nos textos não ficcionais de Pynchon. Ela só aparece uma vez no corpus.

O vocábulo *TIME*, no mesmo corpus, é um Tipo e ele se repete cento e sete vezes. Assim *TIME* é um vocábulo⁸, um Tipo, no dicionário do texto e suas repetições são as Palavras.

Vamos aos números

As duas primeiras medidas estão entrelaçadas inevitavelmente, pois tratam: uma do número de períodos no texto e a outra do tamanho dos períodos. Ou seja, são dois lados da mesma moeda. E fazemos as duas medidas por uma razão didática, porque, às vezes, por questão de objetivos do texto, uma medida pode ser mais convincente, visualmente, do que a outra. Por exemplo: um pesquisador que esteja interessado em demonstrar a fragmentação de um texto, sempre em comparação com outros, vai ser mais convincente se apresentar o número dos períodos nos

⁸ Por uma questão de clareza e desambiguação, todas as vezes que nos referirmos a Tipo(s) e Palavra(s) referentes aos elementos de cálculo, grifaremos com a inicial maiúscula.

corpora. Por isso, mostramos as duas medidas aqui, pois buscamos formar pesquisadores na área de estilometria.

A primeira tabela que apresentaremos, a seguir, mostrará o número de períodos em cada texto e na sequência a tabela com os comprimentos dos períodos.

| | Vineland | Pynonfic | Tinasky | Other_Letters |
|--------------------------------|----------|----------|---------|---------------|
| Número de Períodos nos corpora | 1.075 | 1.170 | 1.397 | 1.685 |

Aqui vemos o tamanho dos períodos nos diversos corpora utilizados medido por número vocábulos.

| | Vineland | Pynonfic | Tinasky | Other_Letters |
|--|----------|----------|---------|---------------|
| Comprimento dos períodos ^{9,10} | 25.81 | 24.43 | 20.45 | 17.09 |

Observem, já nestes cálculos, as discrepâncias entre os textos de Pynchon e os de Tinasky. Podem parecer medidas pouco reveladoras caso não se pause um pouco para pensar sobre elas. Mas, aprofundando um pouco, há de se compreender que essas medidas estão ligadas indissociavelmente à maneira que cada sujeito pontua seu texto e, sabemos, que pontuação varia muito de sujeito para sujeito. A prova disso é que os corpora Vineland (Ficção de Pynchon) e o Pynonfic (Não ficção de Pynchon) estão muito

⁹ **NEOLO** fornece esse número com quinze casas decimais. Utilizamos a regra de arredondamento proposta na ABNT NBR 5891.

¹⁰ Período neste texto considera-se uma estrutura linguística que termine com ponto; ponto de interrogação; reticência; e, ponto de exclamação.

próximos em valores. Mesmo sendo textos com propósitos diferentes, não resta dúvidas da estilística própria que o novelista utiliza.

Nada nos impede de aventar a hipótese de, por Tom Hawkins ser um escritor de ficção em poesia, seu estilo ser mais econômico que o de Pynchon. Dentro da mesma proposta, pensar sobre como esses sujeitos treinados na arte de escrever se diferenciam do homem comum, ordinário, quando observamos as métricas dos escritores em comparação com o corpus de *Other_Letters*, que é um ajuntamento de cartas publicadas juntas com as de Tinasky no mesmo jornal *AVA* e em reação às cartas da *'bag lady'*.

As quatro medidas que apresentaremos na continuidade são medidores de riqueza lexical, calculadas das maneiras mais simples e que já eram usadas no século XIX, quiçá antes, por estudiosos, geralmente, religiosos. São utilizadas fórmulas propostas sem a preocupação com dados que estatisticamente são muito relevantes, destes, principalmente o tamanho dos corpora. Por exemplo: a primeira medida é a contagem simples de hápax no texto. Ora, há de se convir que ao comparar dois corpora de tamanhos diferentes, o corpus maior tem uma grande probabilidade de ter mais hápax. Assim, essa medida só teria eficácia se comparados corpora do mesmo tamanho. No caso de nossa pesquisa, essa medida tem validade porque decidimos que os corpora seriam do mesmo tamanho. Eles variam entre 28.000 e 28.300 Palavras. A pequena diferença no tamanho ocorre pela dificuldade que os algoritmos têm em firmar precisamente esses números. Às vezes, utilizando o mesmo software, as medidas variam de uma pesquisa para outra, mas também porque essa diferença é muito pequena para invalidar os cálculos.

| | Vineland | Pynonfic | Tinasky | Other_Letters |
|-----------------|----------|----------|---------|---------------|
| Número de hápax | 3.420 | 3.496 | 2.980 | 3.352 |

Esses números denotam exatamente o que afirmamos acima: os corpora de Pynchon apresentam um vocabulário mais elástico do que o corpus de Tinasky, talvez pela economia necessária ao poeta Tom Hawkins. Também, o número alto de hápax para o corpus Other_Letters se justifica por se tratar de texto de lavras diferentes, ou seja, esse corpus reúne o vocabulário de dezenas de pessoas, por isto a sua elasticidade.

De qualquer forma, podemos concluir que a coincidência em números dos dois corpora de Pynchon discrepam grosseiramente com o número obtido para Tinasky.

Agruparemos na mesma tabela os próximos dois cálculos por eles seguirem exatamente a mesma lógica. O primeiro divide o número de Tipos pelo número de Palavras (TTR) e o segundo divide o número de Hápax pelo número de Palavras (HTR). Porque serem os números de Hápax e de Tipos sempre menores do que o número de Palavras, o resultado é sempre inferior a um (1) e quanto maior for o resultado, maior a riqueza lexical. Vejamos

| | Vineland | Pynonfic | Tinasky | Other_Letters |
|-----|----------|----------|---------|---------------|
| TTR | 0.195 | 0.195 | 0.173 | 0.191 |
| HTR | 0.123 | 0.122 | 0.104 | 0.116 |

Os números apresentados não devem causar estranheza, visto que eles apenas refletem os dados das tabelas anteriores, sendo que aqui estamos tratando de valores relativos e, portanto, passíveis de revelar qualidade dos corpora: no caso, suas riquezas. Para concluir que os corpora de Pynchon se revelam mais ricos do que os de Tinasky e próximos do corpus `Other_Letters`, isso acontece pelo motivo já mencionado acima: este corpus é formado por textos de diversos autores e cada um deles traz vocábulos diferenciados, o que enriquece o corpus em termos de variedade lexical.

As medidas apresentadas na próxima tabela trazem os números conseguidos com os algoritmos de verificação de riqueza lexical de cálculos mais sofisticados. Eles são: MSTTR, MTL D e o HD-D.

| | Vineland | Pynonfic | Tinasky | Other_Letters |
|-------|----------|----------|---------|---------------|
| MSTTR | 0,717 | 0,723 | 0,718 | 0,723 |
| MTLD | 88,65 | 96,26 | 89,02 | 92,59 |
| HD-D | 0,835 | 0,843 | 0,856 | 0,869 |

A primeira informação importante antes das análises dos números é que, nessas medidas, os números após as vírgulas contam muito: estamos tratando de sintonias finas em que décimos são valiosos. Diferentes das contagens mais toscas que vimos na tabela anterior, onde os decimais tinham pouco valor, aqui são nessas casas após as vírgulas que as diferenças ou similaridades acontecem.

Na medida MSTTR, vale chamar a atenção na similaridade que persiste entre o corpus com textos não ficcionais de Pynchon, quando comparado com o corpus das outras cartas (`Other_Letters`). Eles são os corpora mais

ricos lexicalmente e essa medida referenda as medidas anteriores. Já o Pynchon ficcionista se aproxima muito das cartas de Tinasky. Entretanto, essa medida fica prejudicada porque um corpus é de ficção e o outro de missivas. Mas, a rigor, e para manter a lógica deste texto, a prova oferecida pelo MSTTR ficaria comprometida, pois não ajudaria a alcançar nossos objetivos.

Contudo, continuando as comparações, a situação vira a favor de nossos objetivos quando olhamos os números do MTL. Vejam que os corpora de Pynchon se distanciam mais de 1 (um) ponto do corpus de Tinasky e este discrepa completamente do corpus `Other_Letters`. E as medidas mostradas pelo HD-D confirmam essas diferenças de forma conclusiva. Em verdade, para o leitor mais atento, o que fica claro é que a medida MSTTR é a que mais demonstra a sensibilidade das casas decimais. Porque observados esses décimos e centésimos, todas as medidas nesta medida, fica confirmada a diferença de riqueza dos corpora envolvidos.

Conclusão

Chegamos ao final deste segundo artigo em que tentamos demonstrar, por caminhos da estilometria, que as cartas de Wanda Tinasky não foram escritas pelo ficcionista estadunidense Thomas R. Pynchon. A tarefa já havia sido alcançada antes por artifícios históricos, mas que os pynchomaníacos, especialmente aqueles de fora da academia, frequentemente duvidavam. Ficam para esses mais essas outras evidências que, esperamos, afastem as dúvidas.

No primeiro estudo, já citado referenciado, demonstramos as diferenças entre os textos em questão por via das escolhas lexicais de cada um dos autores, inclusive com demonstração da diferença entre eles na formação de neologismo: Tinasky tinha uma predileção por anglicizar palavras de origem francesas ou alemãs, enquanto Pynchon privilegia as palavras de

origem castelhanas, especialmente vocábulos frequentes no México ou na América Latina. Já neste artigo, recorreremos ao método mais frequente para desambiguar autoria que é o estudo estatístico dos léxicos.

Além de alcançarmos o objetivo primário de demonstrar o estranhamento de Pynchon às cartas de Tinasky, esperamos, com esses dois textos, ter deixado uma trilha metodológica para ser adotada por outros pesquisadores que tenham por missão determinar autoria de textos apócrifos ou anônimos.

REFERÊNCIAS

ANDERSON, B. *The letters of Wanda Tinasky*. San Francisco: Vers Libre Press, 1996.

BRANDÃO, S. C. S. *Neolo*. Disponível em: <https://github.com/joshuacrogey/neolo>. Acesso em: 22 jan. 2016.

BRANDÃO, S. Os porquês de Wanda Tinasky não ser Thomas Pynchon – um estudo de caso (I). *Texto Digital*, v. 14, n. 2, p. 145–155, 2018.

EDER, M. “Does size matter?”, *Digital Scholarship in the Humanities*, London, v. 30, n. 2, p. 167-182, 2015.

FOSTER, Donald. *Author unknown: On the trail of anonymous*. New York: Henry Holt and Company, LLC, 2000.

HALTEREN, H. *et al*. New Machine Learning Methods Demonstrate the existence of Human Stylome. *Journal of Quantitative Linguistics*, 2005. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/09296170500055350>. Acesso em: 22 jan. 2006.

HENDRIX, Jenny. Mistaken Identity. *The Paris Review*, 2012. Disponível em: <https://www.theparisreview.org/blog/2012/01/24/mistaken-identity/>. Acesso em: 10 jun. 2020.

HOLLANDER, Charles. Where is Wanda? The case of a bag lady and Thomas Pynchon, 1997. Disponível em: <http://www.ottosell.de/pynchon/wanda>. Acesso em: 3 mai. 2015.

PENG, R. and Nicolas HENGARTNER. Quantitative analysis of literary styles. *The American Statistician*, v. 56, n. 3, p. 175-185, 2001.

OLSSON, John. *Forensic Linguistics: An introduction to Language, Crime and the Law*. New York: Continuum International Publishing Group, 2008.

PENNEBAKER, J. W.; KING, L. A. Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, 1999. Disponível em: <https://psycnet.apa.org/record/1999-15054-015>. Acesso em: 7 fev. 2021.

PENNEBAKER, J. W. *The secret life of pronouns: What our words say about us*. Nova lorque: Bloomsburry Press, 2011.

TORRUELLA, J.; R. CAPSADA. Lexical Statistics and Typological Structures: A Measure of Lexical Richness. *Procedia*, n. 95, p. 447-454, 2013.

TWEEDIE, Fionda; R. Harald BAAYEN. How Variable Can a Constant Be? Measures of Lexical Richness in Perspective. *Computers and the Humanities*, v. 32, n. 5, p. 323-352, 1998.

NOTAS DE AUTORIA

Saulo Brandão (saulo.brandao@ufrpe.br): Sua primeira graduação foi em Medicina Veterinária pela U.F.R.Pe (1980), posteriormente graduou em Letras pela Faculdade de Ciências Humanas de Olinda (1990), mestrado em Letras pela Universidade Federal de Pernambuco (1995) e doutorado em Letras na mesma universidade (2000), ambos os cursos na área de Teoria da Literatura. Foi pesquisador visitante na Illinois State

University, Bloomington/Normal, a convite do English Department da instituição (1998). Cumpriu missão de pós-doutoramento na UFSC em 2003 e outra na University of Washington - Seattle (set. 2013 - fev. 2015). Atualmente é aposentado como professor Titular da Universidade Federal do Piauí. Trabalha com Línguas e Literaturas Estrangeiras Modernas (Inglês), atuando principalmente nos seguintes períodos: literatura estadunidense pós-1920, literatura inglesa elisabetana e com interesse específico nos seguintes temas: Pynchon e Shakespeare; outros interesses: literatura brasileira de viés fantástico, literatura latino-americana de viés real-maravilhoso, ferramentas telemáticas, estilometria, lexicometria, pós-modernidade e mimetismo. Em outubro de 2020 tomou posse no cargo de Professor Adjunto A na Universidade Federal Rural de Pernambuco, campus em Belo Jardim (UABJ) onde ensina Inglês Instrumental e Metodologia da Pesquisa Científica. Tem as seguintes experiências administrativas: Chefe de Departamento (2001-2002); Coordenador de Pós-Graduação (2004); Coordenador Geral de Pesquisa da UFPI (2005); Presidente da COPEVE (2007-2008); Pró-Reitor de Pesquisa e Pós-Graduação (11/2008 - 1/2013). Em 9/2013 iniciou Estágio Sênior na University of Washington - Seattle e concluiu em 2/2015, com financiamento CAPES. É avaliador de instituições e cursos do BASIS/INEP desde 2006.

Como citar este artigo de acordo com as normas da revista?

BRANDÃO, Saulo. Os porquês de Wanda Tinasky não ser Thomas Pynchon – um estudo de caso (II). *Texto Digital*, Florianópolis, v. 19, n. 1, p. 5-26, 2023.

Contribuição de autoria

Não se aplica.

Financiamento

Não se aplica.

Consentimento de uso de imagem

Não se aplica.

Aprovação de comitê de ética em pesquisa

Não se aplica.

Licença de uso

Este artigo está licenciado sob a Licença Creative Commons CC-BY. Com essa licença você pode compartilhar, adaptar, criar para qualquer fim, desde que atribua a autoria da obra.

Histórico

Recebido em: 13/04/2023.

Aprovado em: 10/07/2023.