TEXTO DIGITAL

Revista de Literatura, Linguística, Educação e Artes

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

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

Saulo Cunha de Serpa Brandão^a

^a Universidade Federal do Piauí, Piauí, Brasil - brandaosaulo@yahoo.com

Palavras-chave:

Thomas Pynchon.
Wanda Tinasky. Neolo.
Cartas.

Resumo: Este trabalho relata uma experiência prática em que utilizo algumas funcionalidades do *software* Neolo para analisar três grupos de textos: 1. Escritos não ficcionais de Thomas Pynchon; 2. As cartas de Wanda Tinasky; e, 3. Cartas que foram publicadas em reação às ideias de Wanda no mesmo jornal. O objetivo é mostrar que é possível comprovar que as cartas de Wanda não foram escritas por Pynchon como criam alguns especialistas da obra do ficcionista. O professor Donald Foster já fez essa comprovação seguindo uma investigação tradicional, histórica. Neste texto demonstrarei como a distinção pode ser feita a partir de análise computacional e atende a métodos quantitativo, qualitativos e aliados a algum dado histórico para fortalecer a proposta. Um último objetivo que almejo com este trabalho é demonstrar que não é preciso fórmulas arrojadas e complexas para fazer o trabalho de distinguir estilo de autores.

Keywords:

Thomas Pynchon.
Wanda Tinasky. Neolo.
Letters.

Abstract: This paper reports a practical experience in which I use some Neolo software function to analyze three groups of texts: 1. Non-fictional writings by Thomas Pynchon; 2. The letters of Wanda Tinasky; and, 3. the Letters that were published in reaction to Wanda's ideas in the same newspaper. The goal is to show that it is possible to prove that Pynchon did not write Wanda's letters as some specialists in the work of the fictionist believed. Professor Donald Foster has already proven this by following a traditional, historical investigation. In this paper, I will demonstrate how the distinction can be made from computational analysis and attends to quantitative, qualitative methods and allied to some historical data to strengthen the proposal. A final goal I aim at with this work is to demonstrate that it takes no bold and complex formulas to do the job of distinguishing authors' style.

Esta obra foi licenciada com uma Licença [Creative Commons Atribuição 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

INTRODUÇÃO

Vou discutir neste texto um experimento feito com a utilização de software de minha autoria que tem como função primária a extração de neologismos e palavras estrangeiras dos textos em análise.

O nome do software é Neolo e além de extrair neologismos, executa mais doze funções; umas poucas eu descreverei rapidamente mais abaixo. As ferramentas foram pensadas para ajudar em pesquisas de determinação de autoria, embora, obviamente, possam ser utilizadas para outras pesquisas a depender da necessidade e criatividade de cada pesquisador.

Eu e minha instituição (UFPI) mantemos o direito de propriedade intelectual sobre *Neolo* (INPI; BR 51 2015 001529-0), mas não temos o menor intuito de comercializá-lo. Ele está disponível em <https://github.com/jcrowgey/neolo>. O algoritmo está escrito em *Python* e os possíveis interessados podem: utilizar, modificar e distribuir livremente. A exigência única que se faz é que nos sejam dados os créditos pela criação original.

Dentre as funções do software, para além da primária anunciada acima, destaco três outras que não são comumente encontradas nos *softwares* que estão disponíveis nos diversos *toolkits* mantidos em várias instituições. Eles são três medidores de riqueza lexical que podem ser utilizados em textos com tamanhos diferentes. Essas ferramentas atendem por acrônimos: MSTTR (Mean Segmental Type-Token Ratio; Johnson, 1944); MTLT (Measure of Textual Lexical Diversity; McCarthy 2005); e, HD-D - os autores não nomearam esse acrônimo, mas subentende-se que se trata de *hypergeometric probability distribution* com mais um D que aparece significando **resultado**. Existe uma diferença que separa a lógica dos dois primeiros do último: nos primeiros o cálculo é feito a partir de seguimentos de léxicos, ou seja, o cálculo parte da separação de séries de palavras seguidas, enquanto o último a escolha das palavras que entram no cálculo é feita a partir de uma seleção por via estatística. Esta ferramenta é muito complexa e exige máquinas muito aquinhoadas para dar conta do trabalho. Como o tempo de processamento é muito longo, sugiro a utilização de computadores dotados de núcleos de processamento da família *Xeon* da *Intel*.

O PROBLEMA

Apresentada brevemente a ferramenta, passo a falar sobre o problema que tentarei resolver utilizando as possibilidades oferecidas por *Neolo*. Trata-se de uma disputa que envolve algumas cartas que apareceram no jornaleco *Anderson Valley Advertise* (AVA), que era publicado no distrito de Mendocino/CA/E.U.A., nos longínquos anos entre 1983-1988. Elas eram assinadas por uma Wanda Tinasky.

Acrescido ao nome da autora cabia ainda mais um detalhe porque ela se autodenominava *the bag lady*. Essa senhora demonstrava ser dona de uma erudição rara, estilo marcado e refinado, sarcasmo agudo, posições políticas libertárias, práticas de contracultura e anti-neoliberais, esquisitas: apesar de apresentar pensamentos libertário sobre a política dos E.U.A., o enfoque na formulação de seus pensamentos políticos é anacrônico: tratava dos assuntos de forma pouco sistematizada, meio caótica, nos moldes da filosofia *hippie* dos anos 1960.

A mesma época do aparecimento das cartas de Wanda Tinasky coincidia com a mudança de Nova York para a Califórnia do ficcionista Thomas Pynchon que tinha ido para o oeste em busca de experiências e características da vida no norte daquele estado. Ele carregava um *blue print* de um novo romance cuja a história seria localizada naquela latitude.

Outro aspecto que a crítica especializada sabia sobre o novo romance era que o novo livro Pynchon mostraria a perspectiva contracultura sob o ponto de vista de *ex-hippies* que com o passar do tempo foram se agrupando na Califórnia por conta da liberdade que lhes era concedida lá. A Califórnia era um estado-santuário¹ antes mesmo de o conceito ser criado.

Então, quer pela localização geográfica de Thomas Pynchon, quer pela temática do livro que era esperado pela crítica especializada no autor, as cartas de Wanda Tinasky foram creditadas a ele por alguns críticos. Os sarcasmo, erudição e posições políticas eram muito parecidas e tratadas estilisticamente aos moldes da escrita do escritor renomado. Na época das cartas e do burburinho feito pelos críticos, Pynchon² negou a autoria dos papeis, através de sua editora Melanie Jackson.

Esse episódio ganhou notoriedade e perdurou na cena cultural dos Estados Unidos por um tempo longo, ao ponto de em 1996 o editor do *AVA* publicar, em livro, as cartas, mas além

¹ Só muito recentemente (2017) a Califórnia foi considerado um estado-santuário. Suas maiores cidades há muito são consideradas cidades-santuários. Quando eu invoco esse termo, santuário, não estou falando exclusivamente sobre a questão migratória, mas sobre as garantias individuais como defendidas pela constituição daquela nação.

² Pynchon é um dos mais reclusos autores da contemporaneidade.

delas, toda as reações por elas provocadas na comunidade de Mendocino e vizinhança e enviadas em forma de cartas endereçadas também ao jornal. Dessa forma resta documentado no livro todo um debate sobre política, arte e cultura da época da publicação no jornal.

A questão que se coloca é: as Cartas teriam sido escritas por Thomas R. Pynchon? A resposta é não! Então a pergunta que movimenta este texto: é possível comprovar linguisticamente que as Cartas não são de Pynchon? Isto vamos discutir e perseguir.

A HISTÓRIA DA REVELAÇÃO

A autoria das cartas já está revelada desde 2000 quando professor Donald Foster publicou o livro *Author Unkown*. No capítulo cinco que tem o título *Wanda, the Fort Bragg Bag Lady*, Foster descreve como ele juntou pistas históricas/geográficas e culturais que o levaram a desconfiar que o autor das cartas era o poeta da *Beat Generation*, Tom Hawkins. Embora o professor tenha feito descobertas importantes, anteriormente, como um soneto de Shakespeare e a autoria do livro *Primary colors*, utilizando ferramentas computacionais para comparar estilos, o caso de Wanda a investigação teve curso tradicional e a confirmação veio quando, com a morte de Tom Hawkins (Wanda), a sua máquina de datilografia comprovou a origem das cartas por conta de falhas em determinados tipos.

ESTA PESQUISA ELEMENTAR – A METODOLOGIA

O trabalho que descreverei na parte quantitativa usa dados puramente linguísticos angariados por ferramenta computacional. Esta que era a tradição nas pesquisas de Foster e que, para ele, nesse caso, se demonstrou insuficiente.

O funcionamento de *Neolo* para pesquisar neologismos é de lógica simples. Trata-se apenas de submeter os textos-objetos da pesquisa a dicionários da língua em que os objetos estão escritos. No caso em tela, o tratamento foi para inglês na modalidade estadunidense. O dicionário que utilizei foi o *The Project Gutenberg Etext of Webster's Unabridged Dictionary* de 1996.³ Como existe interesse que ter o maior número de léxicos possíveis, incluí não só os verbetes, mas todos os conceitos, sinônimos, antônimos, tais como aparece no dicionário. Desta forma contamos com 253.665 verbetes. Neste cômputo estão incluídos

³ O dicionário foi disponibilizado em e-text pelo Gutenberg em 1996, nas ele é o Webster's Unabridged de 1913. O mesmo dicionário de 2013 conta com 600.000 verbetes.

vocábulo ajuntados em uma lista de palavras que, à medida que usamos *Neolo*, descobrimos que são palavras dicionarizadas e que não constam do *Webster*. Exemplo fácil disso são as palavras que flexionam em número de forma regular. Então, por exemplo, existe no dicionário o verbete *girl*, mas não *girls*, por ser um plural regular. Por isto temos listas de palavras que são postas paralelas ao dicionário na hora de utilizar *Neolo* e evitar que o léxico *girls* apareça como neologismo. O mesmo acontece com as flexões dos verbos e variações de gênero, regulares.

Além dos dois corpora-filtros listados acima, existem outros filtros: um com nomes próprios e outro com abreviações e de abreviações dos pronomes de tratamento, que exigem ponto, por exemplo **Mr.** e **Dr.**. A importância desta última lista é para evitar que os pontos utilizados nos pronomes interfiram no cálculo do comprimento médio das sentenças no texto.

Dessa forma cumpre informar que os corpora-filtros crescem permanentemente. Com as adições feitas ao dicionário, hoje ele conta com mais de 400.000 verbetes e quando esse quantitativo é utilizado junto com uma ferramenta de *stemming*⁴ aliado ao *Neolo* a quantidade de verbetes cresce vigorosamente, mas de impossível estabelecimento de tamanho.

Passo a discutir os corpora-objetos: eles são três: 1. Textos das Cartas de Wanda Tinasky; 2. Textos não-ficcionais de Thomas Pynchon; 3. Textos das outras cartas reagindo⁵ às ideias de Wanda (publicadas também no AVA).

Neste trabalho, mostro como a frequência de vocábulos utilizados nos três conjuntos de textos diferem de maneira a matizar as diferenças de autorias convincentemente. Os resultados obtidos dos textos serão comparados com a frequência padrão do uso desses vocábulos por falantes da língua inglesa, conforme demonstrado no *The Frequency Dictionary of Contemporary American English*.⁶ Pelo fato da existência deste dicionário decidi pela não necessidade de utilizar um texto-testemunha na comparação das frequências das palavras escolhidas.

Os textos-objetos têm tamanhos variados, estando na ordem de 26.665 (corpus máximo para os textos de Pynchon não-ficção) até 48.101 palavras. São textos curtos, mas os tamanhos são

⁴ *Stemmings* são ferramentas que são anexadas a programas diversos para identificar palavras a partir de sua raiz lexical.

⁵ Essa reação não tem a conotação negativa. Muitos dos reagentes eram muito favoráveis às ideias de Tinasky.

⁶ Esse dicionário lista as 5.000 palavras de uso mais frequente da língua inglesa.

perfeitamente aceitáveis como populações, porque é suficiente para extrair amostras confiáveis⁷. A regra de ouro na definição das amostras⁸ é: quanto maior a amostra, menor a margem de erro. Assim, explico os casos mais extremos: os textos não-ficcionais de Thomas Pynchon não são amostra, mas população⁹, ou seja: são todos os textos que Pynchon publicou fora os de ficção¹⁰, então não há como dar erro de amostragem para esse corpus; o exemplo menos vantajoso, menos seguro, é o das cartas que reagiram às ideias de Wanda que conta com uma população de 48.101 palavras. Para trabalhar com 99% de certeza de correção dos dados precisaríamos de amostras de 12.334 palavras. Ora, os corpora padronizados contam com 26.665 palavras para se igualarem ao de Pynchon, então não há de se falar em erro por conta da amostragem, porque os corpora-amostragens contam com mais do que o duplo da amostra mínima aceita. Falo de erro possível ínfimo, muito inferior a 1%.

Importante notar que não trabalhamos com *bag of words*. Ou seja, conjuntos de palavras extraídas aleatoriamente dos textos e sem que haja interação sintática entre elas. Os corpora que utilizamos são pedaços estruturados de textos retirados aleatoriamente. Por mais improvável que pareça, o trabalho de determinação de autoria utilizando as palavras mais frequentes, como faremos, é mais preciso utilizando *bag of words* do que usar textos estruturados. Explicando melhor: se o pesquisador trabalha com *bag of words* precisará de um número menor de palavras para determinar autoria do que se lançar mão de textos estruturados. Ressalvando que isso é verdade exclusivamente para se trabalhar com frequência de palavras. Em qualquer dos dois casos, corpora de 5.000 palavras se mostraram suficientes para a tarefa, conforme demonstrou Maciej Eder em seu texto de 2015.

O leitor deve se perguntar o porquê de eu não utilizar os corpora em sua plenitude, ou seja, completo, as populações. A resposta é simples para quem já lida com estudos de corpus, mas não é de tão fácil percepção para neófitos, por isto eu explico: como eu vou trabalhar com a frequência de ocorrências das palavras dentro de um determinado texto as vezes de aparição de uma palavra varia de acordo com o número de palavras da amostra, ou população, podendo modificar assim a sua frequência de ocorrência. Observe a sentença: “O rato roeu a roupa do rei Arthur”. A frequência de ocorrência de qualquer uma das palavras, porque diferentes umas

⁷ Isto nos é ensinado por Eder em um artigo de 2015: Does size matter?

⁸ Os cálculos são feitos utilizando fórmulas matemáticas que parecem ser muito complexas, mas não as são depois que o estudioso entende a lógica por trás das fórmulas. Mas como para estudiosos da área de humanas tudo que envolve número parece impossível, existem máquinas para calcular essas grandezas na internet. Senão, vejam em <http://www.publicacoesdeturismo.com.br/calculoamostral/>

⁹ Esse foi o marco decisivo para o tamanho dos corpora.

¹⁰ Coleta em 2014.

das outras, em caso é de $8/8^{11}$. Note que a simples adição de qualquer outro vocábulo diferente fará que a frequência passe para $9/9$, e se o vocábulo for uma palavra repetida o cálculo mudaria para $8/9$. Por isso a importância de termos textos com tamanhos iguais uma vez que trabalharei com números absolutos.

Existem ferramentas para calcular a frequência relativa dos vocábulos no texto independentemente do tamanho dos textos como já mencionado. No Neolo¹², eu implantei três dessas fórmulas, a saber: MTLT (measure of textual lexical diversity); MSTTR (mean segmental type-token ratio); HD-D (calculation of probabilities using the hypergeometric probability distribution) como já anunciei antes.¹³ Não comentarei com mais vagar porque não intenciono utilizar essas fórmulas no presente trabalho porque foge da proposta aqui desenvolvida que mostrará a diferença de estilo pela frequência das palavras e não pela riqueza lexical e se baseando em ferramentas simples.

Um outro ponto que merece destaque é o do que diz respeito ao tratamento dos textos para utilização em softwares computacionais. No caso de Neolo, o pesquisador tem que ter disponível os corpora em formato “.txt”. Vejam o exemplo das cartas de Wanda e as dos reagentes, que foram os casos mais difíceis; os textos que eu consegui desses corpora foram em PDF (*Portable Document Format*) imagem. Utilizei o aplicativo *Calibre – E-book Management* para fazer a transformação do PDF imagem para o modo “.txt”. Como o programa de transformação não utiliza a ferramenta OCR (*Optical Character Recognition*)¹⁴ ocorrem muitos erros que demandam correção.

Os que trabalham com análise de corpus alocam em seus projetos 60% do tempo da pesquisa exclusivamente para limpeza¹⁵ dos corpora (Larose, 2005). Eu e meus colegas não

¹¹ Devo esclarecer que em tratando de frequência de ocorrência de uma palavra num texto, ou sentença, a forma mais simples de verificar essa frequência é dividindo o número de tipos de palavras pelo número de palavras no conjunto.

¹² Mais sobre Neolo consulte um texto meu publicado na *Texto Digital* em 2017: A Necessidade faz a oportunidade: software Neolo.

¹³ Para evitar explicações que demandariam espaço tempo de produção e de leitura, opto por remeter a um texto que trata dessas fórmulas de maneira mais amigável, didática: Torruella, J.; Capsada R. Lexical Statistics and Tipological Structures: A Measure of Lexical Richness. *Procedia*, No 95 (2013), p. 447-454.

¹⁴ OCR é uma ferramenta comercial e custosa e o nosso grupo de pesquisa opta sempre por aplicativos gratuitos para que os experimentos possam ser replicados por qualquer pessoa que tenha um computador e acesso à internet. Existem máquinas OCR gratuitas na rede, mas sempre trabalham com limites de tamanhos de textos e/ou vezes que as utilizamos.

¹⁵ Extração de um sem número de caracteres não-românicos que surgem no texto após a transformação externos ou internos às palavras. Existem também casos de substituição de caracteres românicos por outros da mesma classe (e.g., é muito comum a substituição do <m>, por <rn> e vice-versa.

conseguimos limpar textos advindos de PDF imagem nos 60% do tempo do projeto. Nosso índice está mais próximo dos 80%, mas com o trabalho que vimos desenvolvendo utilizando comandos do Linux para ajudar, o tempo vem caindo, mas é ainda a maior dificuldade para tocar projetos dessa natureza.

RESULTADOS (1)

Descrita a metodologia utilizada nesse estudo de caso, passemos a observar resultados que são muito impactantes para demonstrar os porquês de Wanda Tinasky não poder ser Thomas Pynchon a partir de elementos linguísticos de estilo.

Abaixo segue uma tabela com a relação de frequência de palavras retiradas da lista de palavras mais utilizadas nos escritos de Thomas Pynchon e que apresentam as maiores discrepâncias entre os textos em comparação. A verificação foi feita a partir das cinquenta palavras mais frequentes¹⁶ nos textos de Pynchon. As numerações referem-se à posição das palavras no ranqueamento de frequência feito por *Neolo*.

Ranqueamento por frequência de uso

Tabela 1

	Pynchon	Tinasky	Outras Cartas	Dicionário de Frequência
Much	46 ^a	179 ^a	141 ^a	298 ^a
Then	47 ^a	393 ^a	108 ^a	77 ^a
Only	49 ^a	116 ^a	100 ^a	333 ^a

Fique claro que citei esses três exemplos por eles serem mais discrepantes e com isso pode-se verificar um forte indício de que autoria das Cartas não pode ser atribuído a Thomas Pynchon. Percebe-se que a frequência do uso das palavras em questão discrepa em três vezes (2) e em duas vezes (1), quando comparada com os escritos de Wanda Tinasky e difere em três vezes (1) e duas vezes (2) quando comparada com as das Outras Cartas. Outra observação que merece destaque é perceber como o uso dessas palavras nos três grupos de cartas diferem da contagem do Dicionário de Frequência. Ou seja, as cartas são muito mais inflacionadas na utilização dessas palavras do que o uso dela pela grande maioria dos usuários da língua

¹⁶ Não foram computadas palavras com carga semântica. Elas são inapropriadas para esse tipo de investigação, uma vez que a frequência delas vai depender do assunto tratado no texto. Utilizamos apenas as palavras funcionais, ou seja, palavras que servem para dar coerência e coesão ao texto. Elas são reveladoras de autoria porque são utilizadas de acordo com a necessidade e foge da questão semântica e estilística. (c.f. PENG, R. e Nicolas HENGARTNER, 2001).

inglesa. No caso de *Then*, Pynchon tem uma utilização diferenciadas das cartas de Tinasky e Outras Cartas e se aproxima mais do que estas do Dicionário de Frequência.

Outras palavras

Como as obras de Thomas Pynchon são meus objetos de pesquisa há muito tempo, sei da esquisitice dele com certas palavras (fora das 50 mais frequentes) e isso me levou a verificar como ficaram essas palavras na lista de palavras com frequências do *Neolo* e o resultado não me surpreendeu, pelo contrário confirmaram minhas suspeitas e pesquisas anteriores, senão vejam:

Tabela 2

	Pynchon	Tinasky	Outras Cartas	Dicionário de Frequência
Yet	188 ^a	539 ^a	734 ^a	277 ^a
Already	271 ^a	671 ^a	738 ^a	339 ^a
Even	60 ^a	100 ^a	98 ^a	498 ^a
Since	185 ^a	189 ^a	137 ^a	997 ^a

A diferença nas discrepâncias de *even* e *since* foram completamente inesperadas. A inflação no uso dessas palavras nos grupos de cartas salta à vista: no caso de *since* falamos de uma utilização mais cinco vezes maior do que a utilização ordinária da palavra. Quando prestamos a atenção em *even* deduz-se que Pynchon utilizou mais de 8 vezes comparada com a frequência utilização mostrada pelo dicionário e nos outros dois grupos, *even* é usado mais de cinco vezes quando os ranqueamentos são dispostos junto ao Dicionário.

O caso de *yet* é peculiar porque os números de Pynchon ficam mais próximos ao número do Dicionário, enquanto as cartas de Tinasky e as Outras Cartas figuram num patamar muito superior ao que encontramos no Dicionário.

Confirmando minha hipótese, demonstro que com a simples comparação da frequência de utilização de palavras funcionais, posso comprovar peculiaridades de cada conjunto de escritos que me permitem inocentar linguisticamente Thomas Pynchon da autoria das cartas de Wanda Tinasky – The Fort Grabb Bag Lady.

RESULTADOS (2)

Neologismos

Analisando os neologismos extraídos por *Neolo*, a primeira constatação que encontramos e que chama a atenção é a utilização de vocábulos estrangeiros e eles formam padrões estáveis para as cartas de Pynchon e as de Tinasky. Os achados no grupo de cartas que reagiram às proposições de Wanda Tinasky não têm valor para tirarmos qualquer conclusão sobre o uso de línguas estrangeiras ou neologismos, uma vez que essas cartas eram escritas por diversos correspondentes e daí a impossibilidade de agrupar para analisar um único padrão.

Essa análise é qualitativa; diferente das mostradas nos quadros de frequência. Esse fato é o ponto nodal de não devermos analisar as Outras Cartas neste quadrante, porque não podemos esperar qualquer padrão sócio-histórico de um corpus que agrega textos de vários autores.

Então, na esteira do que anunciei mais acima, encontrei nas cartas de Pynchon um número de léxicos de origem hispânica, enquanto nas cartas de Wanda o padrão aponta para os mundos francófono e germânico.

Vejam alguns exemplos¹⁷:

Pynchon: cervezas, coños, dassaud, doble, espagnole, l'histoire, narco, negros, parentis, proles, proto, quiero, reflex, sobre, soldat, suite, theologica, verdes.

Tinasky: awfulitch, beaucoup, bruyere, contradictoris, cusan, DACH, destouches, eukmejian, docta, doctore, doubiago, doublago, ecrasez, enmerde, est-ce, gluckschmerz, hausfraus, hukbalahap, katzeff, kiddiekar, lefemme.

RESULTADOS (3)

Fatores extrínsecos – sócio-históricos

As cartas de Pynchon também podem ser identificadas por quem conhece o autor porque elas estão pontilhadas de nomes próprios que estão ligados a Pynchon, quer como amigos do autor ou aqueles que o influenciaram (declarados pelo próprio ficcionista). Dessa forma os nomes como: Fariña, Adams (Henry), Barthelme, Brecht, Carré, Levine, Machiavelli, Mailer, Marquez, Melville, Orwell, Shelley são índices que vinculam os textos que os citam a Pynchon.

¹⁷ As palavras aparecem grifadas tal como estão nos textos.

Vejam que a rede social que envolve Thomas R. Pynchon está revelada nos escritos não-ficcionais do autor, enquanto que os textos de Wanda revelam nomes ligados à política da época e de personagens locais. Nome de pessoas da região onde circulava o *Amderson Valley Advertise* (AVA).

CONCLUINDO

Começo, aqui, a finalizar meu trabalho, mas não sem antes explicar mais um detalhe que merece comentário: fizemos a diferenciação das cartas das três formas: 1. Utilizamos um método quantitativo quando verificamos as cartas pelo viés das frequências diferenciadas para cada grupo de cartas; 2. Fizemos também uma diferenciação pelo uso de palavras em línguas diferentes do inglês – avaliação qualitativa; 3. Por último fiz uso do meu conhecimento histórico da biobibliografia de Thomas Pynchon para mostrar que as cartas do ficcionista trazem nomes de autores e intelectuais que influenciaram ou são amigos dele.

Faço meu parágrafo final para mostrar que as cartas que foram veiculadas pelo *AVA* foram escritas, de fato, por alguém diferente de Thomas R. Pynchon e que muito provavelmente foram escritas pelo poeta Tom Hawkins usando o pseudônimo de Wanda Tynasky conforme muito bem identificado por meios históricos pelo professor Donald Foster.

REFERÊNCIAS

- ANDERSON, B. *The letters of Wanda Tinasky*. San Francisco: Vers Libre Press, 1996.
- EDER, M. “Does size matter?”, *Digital Scholarship in the Humanities*, London, v. 30, n. 2, p. 167-182, 2015.
- BRANDÃO, S. C. S. *Neolo*. Disponível em: <https://github.com/joshuacrogey/neolo>. Acesso em: 22 jan. 2016.
- _____. A necessidade faz a oportunidade: software Neolo. *Texto Digital*, Florianópolis, v. 13, n. 1, p. 124-140, 2017.
- FOSTER, D. *Author unknown: Tales of a Literary Detective*. New York: Henry Holt and Company, 2000.

LAROSE, Daniel T. *Discovering Knowledge in Data: An Introduction to Data Mining*. Hoboken (NY): John Willey & Sons, 2005.

LOVE, H. *Attributing authorship: An introduction*. Cambridge; New York: Cambridge University Press, 2002.

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

SANTOS, Glauber Eduardo de Oliveira. *Cálculo amostral*: calculadora on-line. Disponível em: <http://www.calculoamostral.vai.la> . Acesso em: 23 abr. 2016.

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