

## História Transviada em leitura distante nas palavras-chaves: usando o Voyant Tools para um estudo de caso

Historia Transviada en lectura a distancia en palabras clave: uso de Voyant Tools para un estudio de caso

Ronald CANABARRO

[ronaldcanabarro@gmail.com](mailto:ronaldcanabarro@gmail.com)

Fundação Getúlio Vargas

### RESUMO

O presente texto apresenta um estudo de caso sobre como o uso do Voyant Tools, através de algumas de suas ferramentas, pode nos ajudar na análise de conteúdo dos temas mais abordados em uma parte da historiografia brasileira, por meio das teses e dissertações dos programas brasileiros de pós-graduação avaliados pela CAPES na área de História, entre 1994 e 2020, cujo tema envolveu identidades e questões associadas à dissidências sexuais e desobediências de gênero. Faço uso da nuvem de palavras chaves e dos links de palavras como forma de analisar as temáticas das teses e dissertações. Concluindo que, apesar de simples, são bons elementos para análise de conteúdo, em especial para reconhecimento dos temas e assuntos mais abordados, todavia sugere-se ampliar as pesquisas com uso de ferramentas digitais por parte de historiadores.

### PALAVRAS CHAVE

história transviada; voyant-tools; leitura distante; historiografia

### RESUMEN

Este artículo presenta un estudio de caso sobre cómo el uso de Voyant Tools, a través de algunas de sus herramientas, puede ayudarnos en el análisis de contenido de los temas más abordados en una parte de la historiografía brasileña, a través de las tesis y disertaciones de los programas brasileños de postgrado evaluados por la CAPES en el área de Historia, entre 1994 y 2020, cuya temática involucró identidades y preguntas asociadas a la disidencia sexual y la desobediencia de género. Utilizo la nube de palabras clave y enlaces de palabras como una forma de analizar los temas de tesis de maestría y doctorado. Concluyendo que, aunque simples, son buenos elementos para el análisis de contenidos, especialmente para el reconocimiento de los temas y temas más abordados, sin embargo, se sugiere ampliar la investigación utilizando herramientas digitales por parte de los historiadores.

### PALABRAS CLAVE

historia Transviada; voyant-tools; lectura a distancia; historiografía

## 1. INTRODUÇÃO

A presente resenha visa apresentar um estudo de caso de leitura distante através do uso do Voyant Tools, um ambiente web de leitura e análise de textos digitais<sup>1</sup>, que teve sua tradução para o português por Fernando Marineli. O *corpus* de análise é parte da minha tese de doutorado em andamento em História, Política e Bens Culturais, pela Fundação Getúlio Vargas (Rio de Janeiro, Brasil).

As fontes para essa pesquisa se centram nas dissertações e teses defendidas nos programas de pós-graduação no Brasil, avaliados pela Capes - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, na área de História, que versam especificamente sobre as dissidências sexuais e desobediências de gênero, e fazem parte do Catálogo de teses e dissertações da Capes<sup>2</sup>.

O Brasil possui hoje um total de 129 cursos de pós-graduação *stricto sensu* em História, sendo 70 mestrados acadêmicos, 12 mestrados profissionais, 44 doutorados acadêmicos e 3 doutorados profissionais, pertencentes a 82 programas de pós-graduação, em 72 instituições de ensino superior (IES), segundo dados da Capes<sup>3</sup>.

Para localizar as teses e dissertações que versem sobre a temática construí uma lista de palavras na intenção de abarcar a maioria das definições e termos êmicos utilizados para se referir a pessoas dissidentes sexuais e desobedientes de gênero. A construção dessa lista de palavras se deu a partir de pesquisas de autoras da biblioteconomia (Lara & Silva, 2004), historiadoras/es (Rodrigues, 2012; Gomes, 2018) e antropólogos (Green, 2000a e 2000b; Mott, 1994).

As palavras que elenquei para as buscas, uma a uma, a partir da pesquisa livre por palavras, são essas: abjeto, abjeção, aids, anormal, anormais, bagaxa, bdsm, bicha, bixa, bissexual, bissexualidade, bissexualismo, corpo, fanchono, fanchona, fresco, gay, guei, gênero, glbt, gls, hermafrodita, hermafrodito, hermafroditismo, hiv, homoafetividade, homoafetivo, homoerotismo, homofobia, homossexuais, homossexual, homossexualismo, homossexualidade, inversão sexual, intersexual, intersexualidade, intersexualismo, invertido sexual, lesb\*, lesbianas, lesbiandade, lesbianismo, lésbica, lesbofobia, lgbt, masculinidade, mhb, movimento homossexual brasileiro, michê, não-binário, pederasta, pederastia, pederast\*, pervertid\*, prostituição, prostituta, prostituto, prostitu\*, prostituição masculina, puta, puto, queer, sadomasoquismo, sexo, sexual, sexual\*, sexualidade, sífilis, sífilítico, sodomia, sodomita, sodom\*, safismo, sapatão, terceiro sexo, transexual, transexualidade, transexualismo, transfobia, transgênero, transgeneridade, transviado, transviada, transfobia, travesti, travestismo, tribadismo, uranismo, ursos.

O total de termos pesquisados foi de 89. Chegando assim a um montante de 4.465 (quatro mil quatrocentas e sessenta e cinco) dentre as teses e dissertações que trazem, pelo menos uma

---

<sup>1</sup> Disponível em: <https://voyant-tools.org/>.

<sup>2</sup> Disponível em: <http://catalogodeteses.capes.gov.br/>.

<sup>3</sup> Disponível em: <https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/programa/quantitativos/quantitativoAreaConhecimento.jsf?areaAvaliacao=40>.

vez, alguma das palavras que se relacione com as da lista. Ou seja, um percentual de 17,91% do total de teses e dissertações da área de avaliação em História no Brasil.

Todavia, esse número não indica que todos esses trabalhos abordem a temática que proponho analisar. Desse total foi necessário avaliar os títulos um a um e, caso trouxesse alguma das palavras da minha lista específica, ele era separado para um segundo momento de análise. A escolha de levantar as pesquisas a partir do título foi baseada na lógica da maioria dos autores da área de metodologia que indicam que o título deve ser objetivo, mas chamativo, trazendo as informações essenciais para o leitor e podendo dar uma ideia do que se trata a pesquisa (Laville & Dione, 1999; Henry-Silva, Soeiro & Camargo, 2009)

Dentro dessa observação dos títulos eu incluí, ainda, as palavras que fazem uma relação de saúde/doença na sua temática e que historicamente se interseccionam nas construções das identidades de dissidentes sexuais e/ou de gênero, como hiv/aids, sífilis e tuberculose, pensando a partir da ideia de “doença como metáfora” (Sontag, 2009). Foram também consideradas todas as atividades de trabalho sexual, entendidos na forma de todo e qualquer local ou atividade profissional cuja prática envolva a oferta de serviços sexuais.

Incluí ainda para análise todas as dissertações e teses que traziam em seu título um nome próprio, assim, essa pesquisa era colocada numa lista para segunda análise. A inclusão de nomes próprios se deu porque existe um número de pesquisas que focam em pessoas que viveram publicamente como dissidentes sexuais e/ou de gênero, tais como Caio Fernando Abreu, Herbert de Souza, Cassandra Rios, entre tantas outras.

O intuito foi abarcar o máximo possível de dissertações e teses que tratam especificamente da temática de gênero e sexualidade dissidentes. Ao final dessas etapas cheguei a um corpus de 151 pesquisas que abordam especificamente sobre sexualidades e identidades dissidentes da cisheteronormatividade<sup>4</sup>. Ou, como podemos identificar, um coeficiente de 1,61% do total de teses e dissertações defendidas na área de pós-graduação em História, no Brasil, que abordam especificamente as dissidências sexuais e desobediências de gênero.

Essas pesquisas passaram a compor um banco de dados em forma de tabela eletrônica, formato/extensão XLSX, que descreve todos os metadados principais dos trabalhos, como título, autoria, orientação, universidade, ano de defesa, resumo, entre outros. As pesquisas foram copiadas em sua versão digital (aquelas já nascidas digitais), ou em sua versão digitalizada, feitas em formato físico no papel e posteriormente digitalizadas para que se tivesse acesso através de visualização digital. Todas as localizadas em formato digital ou digitalizado foram copiadas para o

---

<sup>4</sup> Flexionando os conceitos de cisheteronormatividade e heteronormatividade. Para saber mais sobre os conceitos ver ROSA, (2020).

meu computador e para minha nuvem.

## 2. AS PALAVRAS-CHAVES COMO INDICADORES DOS CONTEÚDOS

Para iniciar minha análise, escolhi olhar para todas as palavras-chaves das teses e dissertações que compõe o meu corpus. As palavras-chave foram extraídas e inseridas na tabela com os metadados de cada uma. Vale lembrar que essas informações fazem parte da elaboração dos dados chamados de pré-textuais, ou seja, aqueles compreendidos entre a capa e o sumário de uma dissertação ou tese.

A prática do uso de palavras-chaves é antiga e tem como principal objetivo servir, junto com o Resumo e suas versões em inglês –as keywords e o abstract–, de auxiliares nos indexadores e facilitar a pesquisa por determinada palavra e/ou termo. De um modo geral as palavras-chaves indicam quais os principais assuntos abordados naquela pesquisa (Henry-Silva, Soeiro & Camargo, 2009).

Para fazer a leitura distante do corpus utilizei a plataforma Voyant Tools. Segundo consta no próprio sítio da internet: “Voyant Tools é um ambiente web de leitura e análise de textos digitais”, e sua tradução para o português foi feita por Fernando Marineli. É uma ferramenta gratuita e bastante completa que permite que o usuário faça análises de textos sem precisar baixar um programa específico em seu computador, bastando acessar o site. Feito em código aberto permite aos usuários e programadores aperfeiçoarem e, até mesmo, proporem novas ferramentas.

Em seu navegador de preferência, a primeira tela apresentada será uma janela bastante simples e intuitiva para que você inicie sua análise, conforme imagem abaixo:

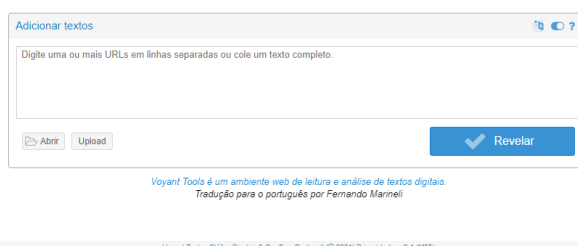


Figura 1. Captura de tela da página do Voyant Tools.

Em sua seção ajuda das ferramentas, em especial no Creating a corpus (criando um corpus), há uma explicação detalhada de quais os passos a serem seguidos e as opções de formatos de corpus de pesquisa aceitos, indo da simples colagem na caixa de texto no navegador, a opção de abrir um corpus já existente, fazer um envio (upload) de um ou mais arquivos existentes no seu computador, entre outras possibilidades de formatos em tabelas do excel (extensões xls ou xlsx), arquivos de internet (extensão html), XML-TEI, Jason, entre outros.

Um apontamento importante é que os dados enviados para análise ficam públicos. Toda-

via, o sistema gera um link aleatório que dificilmente será encontrado por alguém que não tenha o endereço completo. Todavia, caso você esteja trabalhando com dados sigilosos, você pode inserir uma senha antes da geração de seu *corpus* ou mesmo utilizar uma versão offline disponível no GitHub.

Para a minha análise extrai e montei uma lista com todas as palavras chaves das 151 teses e dissertações levantadas através da busca por palavras no catálogo de teses e dissertações da Capes, através dos critérios de inclusão e exclusão já explicitados anteriormente. O arquivo inicial, sem a limpeza dos dados, ficou composto por 1155 (um mil cento e cinquenta e cinco) palavras.

Toda análise exige critérios de exclusão de palavras que não são importantes para os objetivos de cada pesquisa. Esse processo é importante para remover possíveis inconsistências, como palavras com grafia errada ou exibida de formas diferentes. Serve também para padronizar maiúsculas ou minúsculas, palavras com hífen e acentuação gráfica, entre outras coisas. Ainda, é preciso observar preposições, artigos e conjunções (Lemos, 2013, p. 8).

Num primeiro momento padronizei as palavras deixando-as todas em letras minúsculas. Após, tornei todas as palavras que tinham alguma repetição na sua forma singular e plural, como por exemplo “travestis” e “travesti”, optando por deixar todas no singular. Além disso, excluí os espaços, hifens e barras entre as palavras e ou termos compostos, fazendo com que as palavras-chaves, como por exemplo “Jornal Lampião da Esquina” ou “homo-afetividade”, passassem a ser analisadas e comparadas como uma única palavra - “jornallampiãodaesquina” e “homoafetividade” -, essa escolha me permitiu olhar para cada palavra-chave inserida, de forma única e sem interação direta entre elas. Após essa padronização, o documento passou a ter 694 (seiscentos e noventa e quatro) palavras. Esse processo é chamado, muitas vezes, de limpeza dos dados.

Por fim, inseri uma lista de stopwords. O próprio sistema do Voyant têm uma lista de stopwords elaborada no qual você pode selecionar e excluir ou incluir palavras. No meu caso, incluí, além das preposições, artigos, pronomes, numerais ordinários e verbos de ligação, a palavra “história”, uma vez que ela sozinha não me traz nenhuma novidade, sendo que todo o *corpus* se refere a pesquisas na área de História.

Feito isso iniciei a análise na plataforma que você pode acessar pelo endereço <https://voyant-tools.org/?corpus=dbfe15fc48ab2f6d7bd98c4f81d3ce6d> e que ficará visualizável conforme sinalizado abaixo.

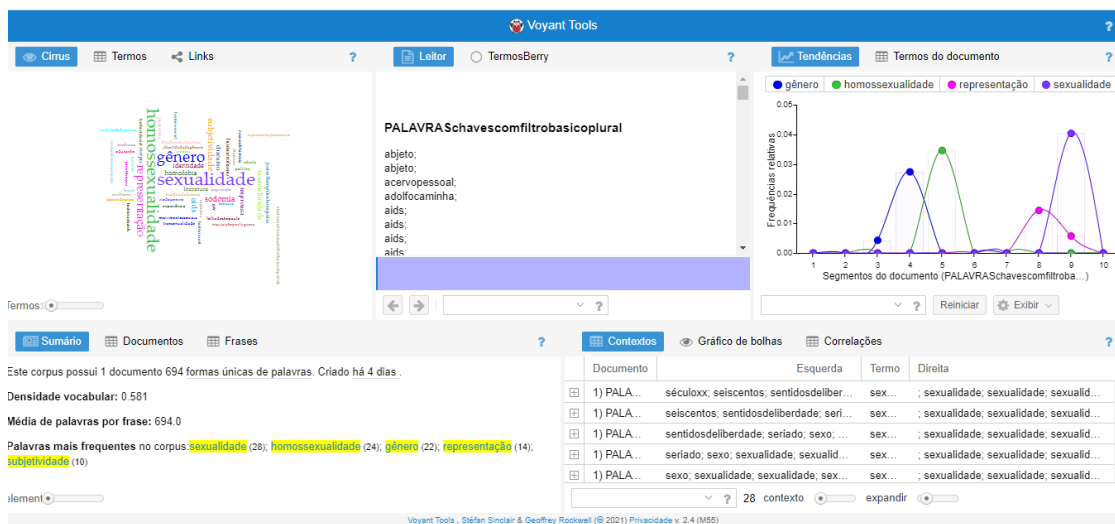


Figura 2. Captura de tela da análise do corpus de palavras-chave no Voyant Tools.

Um ponto observado já de início é que o Voyant Tools não salva suas inserções de palavras diretamente nas stopwords já existentes, assim, toda vez que alguém abrir o link acima a palavra “história” surgirá entre as mais citadas, tendo que ser inserida novamente, caso queira seguir a análise que fiz aqui. Para tal inserção é necessário selecionar ao lado do ponto de interrogação logo acima da nuvem de palavras onde é exibida o “definir as opções para essa ferramenta”, conforme sinalizado abaixo.

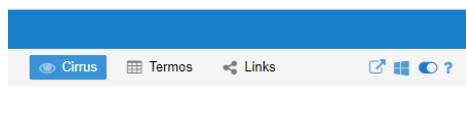


Figura 3. Captura de tela para demonstração de onde alterar a lista de stopwords.

Por padrão o Voyant Tools retorna as 55 palavras mais citadas do corpus para apresentar a nuvem de palavras. Caso você queira aumentar ou diminuir esse quantitativo basta arrastar para um lado ou para outro a barra de termos, conforme indicado na figura abaixo.

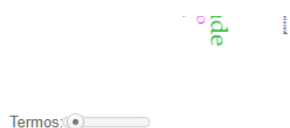


Figura 4. Fragmento de captura de tela do Voyant Tools.

Aqui vou me centrar nas 5 palavras mais citadas e que ganham destaque na nuvem de palavras. Um primeiro destaque é para a palavra “sexualidade”, indicando que na área das dissidências sexuais e de gênero a ênfase nos estudos de sexualidade humana são preponderantes. É importante observar que apesar de bastante genérica a repetição do termo pode indicar que as pesquisas tem avançado sobre a sexualidade na história e a partir da própria história da sexualidade como um campo de pesquisa ou, pelo menos, como uma área de análise dentro da História.

A frequência da palavra “homossexualidade”, em 24 repetições, nos indica um maior número de pesquisas sobre a homossexualidade, em especial a masculina, em detrimento de outras orientações sexuais ou identidades de gênero. Esse é um dos pontos a se pensar dentro da área: a

da representação da homossexualidade masculina como um dos assuntos mais abordados. Essa mesma situação já foi observada em outros campos do saber, como na psicologia (Gaspodini & Falcke, 2019), por exemplo.

Desde a problematização de Joan Scott (1990) sobre o “gênero como uma categoria útil para análise histórica”, que a área de estudos vêm se ampliando, trazendo novos matizes de análises, adaptando novos métodos e abrindo caminhos para repensar a própria história. As problematizações de gênero abriram caminho para que os estudos sobre sexualidade também passassem a ser validados e reconhecidos dentro da academia.

A presença de termos como “subjetividade” e “representação”, como forma de dar visibilidade à história voltada para as representações sociais, aparecem de forma bastante importante e demonstram, também, as possíveis intersecções e aproximações entre a Psicologia e a História.

Seria possível ampliar em muito essa análise observando os demais termos que aparecem e suas respectivas repetições, todavia para fins dessa resenha não ampliarei essa discussão.

Ainda na ferramenta Cirrus encontramos uma análise de links de palavras. Indicando como determinadas palavras se unem a outras através da composição das frases no corpus. Demonstra como determinados termos estão mais diretamente ligados a outros, e, por conseguinte, nos permite fazer uma leitura um pouco mais aprofundada da temática.

As três principais palavras centrais nos links são: homossexualidade, gênero e sexualidade. Elas se encontram com outras doze palavras-chaves que nos trazem indícios de questões e temas abordados nas pesquisas: homofobia, homossexualidade masculina, homossexualidade afeminada, sexo, gay, homossexuais, gmagazine, geração mariel, gênero e masculinidade no século XIX, seriado, sexualidade aspectos históricos e sexualidade história, conforme imagem abaixo.

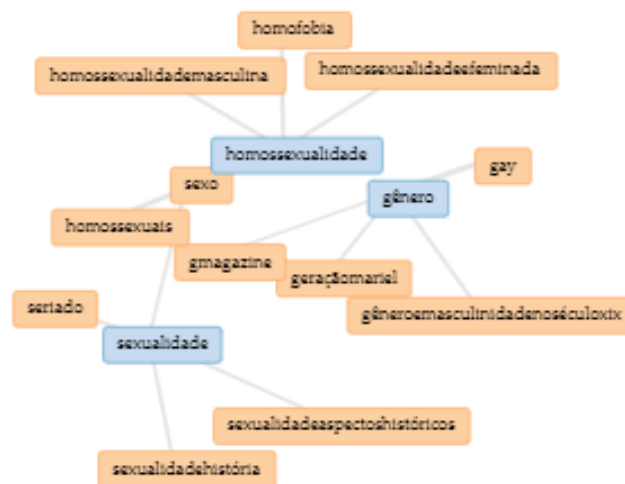


Figura 5. Captura de tela Voyant Tools.

Ao observarmos os termos que se ligam a palavra homossexualidade: homofobia, homossexualidade masculina e homossexualidade afeminada, podemos trazer inferências do quanto a questão da violência, e o estudo sobre ela, estão diretamente ligada à orientação sexual. Ainda, nos apresenta uma outra questão que é a das construções das masculinidades, pensando aqui os

processo de homossexualidade masculina e homossexualidade afeminada. Ambas as situações entram em análises e disputas sobre os atravessamentos com o machismo, a heteronormatividade e a própria misoginia.



Figura 6. Captura de tela Voyant Tools.

Quando olhamos para o termo “gênero” encontramos ligações diretas com: “gay”, “gmagazine”, “geração Mariel”, “gênero e masculinidade no século XIX”. As duas primeiras palavras interligadas à gênero são “gay” e “G Magazine”. Essa última referindo-se a revista de conteúdo erótico gay e que existiu entre 1997 e 2013, foi sucesso, principalmente, entre gays e como imprensa alternativa (Costa, 2018). Mas nos traz um outro ponto, qual seja: a incorporação de novas fontes de pesquisa como revistas eróticas e com muitas imagens tidas durante muito tempo como apenas pornográficas. Já a palavra gay reforça o que já foi dito sobre a presença de uma maioria de pesquisas que abordam a homossexualidade masculina.



Figura 7. Captura de tela Voyant Tools.

O último termo a se observar nos links é “sexualidade”, ligada diretamente a: “sexo”,



“seriado”, “sexualidade aspectos históricos”, “sexualidade história”. No caso do primeiro termo, “sexo”, em conjunto com “sexualidade aspectos históricos”, “sexualidade história”, nos reforça sobre os estudos de história da sexualidade presente na área. Outro ponto que pode nos indicar é a ideia de indivíduo sexuado, nesse caso o sexo representando as identidades sexuais dos indivíduos. Mais uma vez a questão das novas fontes e das novas abordagens se faz presente, e se afirma como uma das características da área. O termo “seriado” nos remete às análises que envolvem o cinema e outros tipos de audiovisuais como fontes para a pesquisa da/na área da História.



Figura 8. Captura de tela Voyant Tools.

### 3. CONSIDERAÇÕES FINAIS

Voyant Tools, enquanto “ambiente web de leitura e análise de textos digitais”, têm ainda vários outros tipos possíveis de ferramentas, incluindo termos Berry, gráfico de tendência das palavras, termos do documento, um sumário com cálculos da densidade dos vocábulos e média de palavras por frase, por exemplo, além de uma breve análise de conteúdo, com correlações e gráfico de bolhas. Todas essas ferramentas nos ajudam a fazer a análise na leitura distante, em especial em corpus bastante extensos. Todavia, nesse estudo de caso, me centrei na ferramenta Cirrus, descrita nos parágrafos anteriores.

Por fim, observo que a nuvem de palavras e a análise de links se mostram ferramentas úteis para identificar insights sobre as análises de conteúdo mais simples, além de permitir que se estabeleça uma forma de visualização bastante lúdica dos aspectos de um determinado texto, ou corpus de textos, deixando a palavra ou termo relevante em destaque e, também, com seus links diretos.

Considero, ainda, necessário construirmos aproximações entre as áreas de ciência de dados, Computação e História, para a produção conjunta de algoritmos de mineração de dados e, também, para a disponibilização de mais programas e sítios de internet que permitam aos usuários maiores possibilidades de escolha.

**REFERÊNCIAS BIBLIOGRÁFICAS**

- Costa, G. D. (2018). Prazeres revelados: produção de sentidos de consumo e subjetividade na Revista G Magazine. Mestrado em História. Universidade Estadual de Montes Claros.
- Gaspodini, I. & Falcke, D. (2019). Estudos psicológicos brasileiros sobre preconceito contra diversidade sexual e de gênero. *Estudos Interdisciplinares em Psicologia*, 10, 59-79.
- Green, J. N. (2000a). Além do carnaval: a homossexualidade masculina no Brasil do século XX. São Paulo: UNESP.
- Green, J. N. (2000b). “Mais amor e mais tesão”: a construção de um movimento brasileiro de gays, lésbicas e travestis. *Cadernos Pagu*, 15, 271-295.
- Gomes Junior, J. (2019). Sobre frescos e bagaxas: uma história social do homoerotismo e da prostituição masculina no Rio de Janeiro entre 1890 e 1938. Mestrado em História. Universidade Federal Fluminense.
- Henry-Silva, G. G., Soeiro, M. I. P. & Camargo, A. F. M. (2009). Caminhos para a publicação científica. *Boletim da Sociedade Brasileira de Limnologia*, 37, 1-10.
- Lara, M. L. G. & Silva, C. R. (2004). Os termos relativos ao segmento GLBT (Gays, Lésbicas, Bissexuais e Transgêneros) no contexto das linguagens documentárias. *Informação & Informação*, 9(1/2), 37-47.
- Laville, C. & Dionne, J (1999). A construção do saber: manual de metodologia da pesquisa em ciências humanas. Porto Alegre: Artmed.
- Lemos, E. dos S. (2013). A aprendizagem significativa: estratégias facilitadoras e avaliação. Periódico do Programa de Pós-Graduação em Educação da UCDB (Série Estudos), 21.
- Mott, L. (1994). Etno-História da homossexualidade na América Latina. Seminário Taller de História de las Mentalidades y los Imaginarios. Pontificia Universidad Javerina de Bogotá, Colômbia, Departamento de História e Geografia.
- Rodrigues, R. C. C. (2012). De Daniele a Chrysóstomo: quando travestis, bonecas e homossexuais entram em cena. Doutorado em História. Universidade Federal Fluminense.
- Rosa, E. B. P. R. (2020) Cisheteronormatividade como instituição total. *Cadernos PET*, 18 (2), 59-103.
- Scott, J. (1990). Gênero: uma categoria útil de análise histórica. IN: Mulher e realidade: mulher e educação. *Vozes*, 16(2).
- Sontag, S. (2007). Doença como Metáfora / Aids e suas Metáforas. São Paulo: Cia. das Letras.