

Análise do discurso digital: sobre o arquivo e a constituição do *corpus*

(Analyse du discours numérique: a propos de la constitution du corpus)

Cristiane Dias¹

¹ Laboratório de Estudos Urbanos – Universidade Estadual de Campinas (Labeurb/ Unicamp)

crisdias@unicamp.br

Résumé: L'objectif de cet article c'est de proposer une réflexion, de la perspective de l'analyse du discours, a propos de la constitution du corpus de recherche de l'analyse du numérique. À la fois, on va considerer des notions déjà connus au champ théorique de l'analyse du discours, comme celle de "découpage", "archive", "circulation".

Mots-clés: archive; discours numérique; internet; corpus

Resumo: O objetivo desse artigo é propor uma reflexão, da perspectiva da análise de discurso, sobre a constituição do corpus de análise do digital. Para tanto, vamos considerar noções já conhecidas no próprio campo teórico da análise de discurso, como "recorte", "arquivo", "circulação".

Palavras-chave: arquivo; discurso digital; internet; corpus.

Introdução

A constituição do *corpus* em análise de discurso, desde o início dessa teoria, foi uma questão posta em discussão. A preocupação central dessa disciplina de entremeio sempre foi, justamente, não tomar a constituição do *corpus* separada das condições de produção, recortando amostras de representatividade, seja da língua, por meio de equivalências gramaticais, seja do discurso, por meio de comparação de enunciados. Guilhaumou e Maldidier (1994) explicitaram essa problemática no texto *Effets de l'archive*, apontando para o arquivo como a questão que permitiu a eles dar uma resposta ao problema do *corpus*.

Essa questão retorna, hoje, uma vez que muitos têm sido os trabalhos em análise de discurso que se dedicam à análise do discurso digital, constituindo, assim, seus *corpora* de trabalho a partir da própria rede internet ou de outros dispositivos a ela conectados.

Esse artigo tem como objetivo, portanto, refletir, da perspectiva da análise de discurso, sobre como se dá a constituição do *corpus* de análise do digital, levando em conta noções já conhecidas no próprio campo teórico da análise de discurso, como "recorte", "arquivo", "circulação".

Do arquivo na análise de discurso

A questão do arquivo em análise de discurso tem a ver com duas outras questões teóricas fundamentais: a leitura e a constituição do *corpus* (GUILLAUMOU; MALDIDIER, 1994). No que concerne à questão do *corpus*, podemos dizer que se trata de um conjunto de formulações produzido pelo próprio processo de interpretação do discurso, no confronto com o arquivo.

Ao falarmos de arquivo, não o consideramos como um acúmulo de documentos organizados e chancelados por uma instituição ou área de conhecimento. Não se trata do arquivo como absoluto de materiais ou campo de documentos pertinentes sobre uma questão. Falamos, sim, da “materialidade do arquivo”, que, segundo Guilhaumou e Maldidier (1994, p. 92), “impõe sua própria lei à descrição”. Isso porque considerar o arquivo em sua materialidade implica encontrar na prática de análise de discurso o momento da interpretação, em relação ao da descrição, num batimento entre um e outro (PÊCHEUX, 2008, p. 54). E disso resulta a própria constituição do *corpus*. A materialidade do arquivo, portanto, é aquilo que faz com que ele signifique de um modo e não de outro, que faz com que ao se deparar com ele, o sujeito o recorte de maneira x e não y. Um mesmo arquivo nunca é o mesmo, por causa da sua materialidade.

Assim, a “materialidade descritível” do arquivo: uma palavra, um enunciado, uma imagem, um gesto, uma letra, uma cor, etc. “coloca em jogo o discurso-outro como espaço virtual de leitura” (PÊCHEUX, 2008, p. 55), marcando, do interior da materialidade, a relação com a memória histórica. Se a descrição instala o real da língua (equivoco, falha, eclipse, etc.), a interpretação instala o real da história (contradição),¹ um alternando-se ao outro na própria complexidade do arquivo, tomado em redes de memória, pondo assim em relação língua e história.

O funcionamento do arquivo é opaco, salientam Guilhaumou e Maldidier (1994). Essa opacidade se impõe, para eles, no momento em que a análise do *corpus* se torna insuficiente para mostrar a relação língua e arquivo e discurso e arquivo – no caso, o discurso revolucionário – e eles sentem a necessidade do “retorno ao arquivo”: “a consideração da dispersão máxima do arquivo torna complexo o procedimento do analista de discurso” (GUILHAUMOU; MALDIDIER, 1994, p. 106). É essa exploração complexa do arquivo que mergulha o pesquisador na materialidade dos sentidos. Portanto, para Guilhaumou e Maldidier (1994), qualquer identificação puramente institucional do arquivo é insuficiente, pois diz pouco sobre o funcionamento dos discursos.

Desse modo, ao constituir um determinado *corpus* a respeito de uma questão, ou em torno de um acontecimento, o fizemos por meio da construção de uma “unidade discursiva” (ORLANDI, 1984) ou recorte de formulações feitas em certas condições de produção, que levam em conta a linguagem e a situação.

É nesse ponto que podemos trazer a questão da leitura como constitutiva daquela do arquivo. Em geral, como adverte Pêcheux (1981), quando se trata do arquivo, contornamos a questão da leitura, questão esta cara à Análise de Discurso. Mas não se trata tampouco da leitura literal, trata-se, ao contrário, de “adicionar sistematicamente a leitura à fragmentação espontânea das sequências para liberar completamente a matéria verbal [ou não verbal] dos restos de sentido que ainda a aderem [...]”² (PÊCHEUX, 1981, p. 16). Em outros termos, trata-se de tirar a leitura de qualquer relação com a evidência.

Nesse sentido, o memorável texto de Pêcheux (2010) “Ler o arquivo hoje”, muito nos ensina, pois vai por em relação o arquivo e a leitura, na contramão de uma tentativa

¹ No caso da análise de Guilhaumou e Maldidier (1994), trata-se da divisão do processo revolucionário contra si mesmo.

² Tradução livre de: “rajouter systématiquement sur le morcellement spontané des séquences, pour achever de libérer la matière verbale des restes de sens qui y adhèrent encore [...]”

de regulação da leitura, de gerenciamento da memória coletiva, de “gestão administrativa dos documentos”, enfim, do desenvolvimento de “*métodos de tratamento em massa*” do arquivo textual” (PÊCHEUX, 2010, p. 52), supondo “torná-los facilmente comunicáveis, transmissíveis e reproduzíveis”. O que Pêcheux mostra é que, ao contrário dessa uniformização da leitura de arquivo, há gestos de leitura diferenciais e contraditórios na construção do arquivo. O autor reivindica “o reconhecimento da materialidade da língua como [constitutiva do] incontornável do pensamento” (PÊCHEUX, 2010, p. 57). E o que ele propõe é a constituição de um

[...] *espaço polêmico das maneiras de ler*⁴, uma descrição do ‘trabalho do arquivo enquanto relação do arquivo com ele-mesmo, em uma série de conjunturas, trabalho da memória histórica em perpétuo confronto consigo mesma. (PÊCHEUX, 2010, p. 51)

Pondo em relação esse *espaço polêmico das maneiras de ler* e a leitura de arquivo nos tempos de internet, deixo aqui uma questão: não seríamos nós o “conjunto de todos os outros”, referido por Pêcheux ao falar da divisão entre os clérigos, na idade média, por meio da qual alguns eram “portadores de leitura e de uma obra própria” e outros, “o conjunto de todos os outros”, exerciam uma leitura que impunha ao sujeito-leitor um apagamento, uma vez que se dava por meio da repetição de gestos como “cópia, transcrição, classificação, indexação, codificação, etc.”? Afinal, não operamos nós, hoje, frequentemente, na relação com o arquivo digital, com gestos da mesma ordem: copiar, colar, compartilhar, recortar, extrair...? Contudo, em vez dos clérigos, temos, hoje, a serviço do Estado e do empresariado, a grande mídia... (mas também a escola, a ciência...).

Arquivo e materialidade digital

Com essa questão, entro numa segunda parte desse artigo, que é refletir sobre o arquivo digital, e que tem a ver, ainda recorrendo a Pêcheux (2010, p. 54), “com a relação da nossa sociedade com sua própria memória histórica”. Quais as consequências das formas de leitura de arquivo que se estabelecem com o digital, sobre a relação da sociedade com sua memória histórica?

Em face da materialidade digital, como fica a questão do arquivo e com ela a da leitura e a da constituição do *corpus*?

Ao citar Canguilhem sobre a questão sentido, Pêcheux (2010, p. 58) mostra que as máquinas podem produzir conexões entre os dados, mas os dados fornecidos pela máquina ao sujeito não estão em relação com o que este se propõe a partir deles. No caso de um buscador, como o Google, por exemplo, é preciso atentar para isso e não subestimar o “fato da língua” na leitura do arquivo. Em outros termos, não tomar como uma evidência do arquivo o resultado da busca, porque ela não é mais do que dados em relação algorítmica numa memória metálica. É preciso, no entanto, atentar para as correspondências que esses “dados” engendram em nós, o que já se dá a partir de uma filiação à memória histórica, de um trabalho do arquivo. É isso considerar o digital em sua materialidade.

³ Grifos do autor.

⁴ Grifos do autor.

No que diz respeito aos arquivos digitais, falamos, da perspectiva de Paveau (2014) ao tratar do “discurso digital” ou “discours numérique”, de materiais produzidos na e para a internet. Quanto a eles, é preciso atentar para algumas características:

– *temporalidade*: constituída por outros paradigmas que escapam a qualquer cronologia. O tempo do digital é o do acesso e da circulação. Um arquivo digital é sempre atual ou, melhor dizendo, passível de atualização pelo acesso. (Ex.: ao comentar uma postagem do Facebook, independentemente da data em que foi postada, ela se atualiza na linha do tempo).

– *instabilidade do arquivo*: diz respeito à sua mutabilidade. É comum lidarmos com *sites*, textos, *blogs*, vídeos, etc. que sofrem atualização ou ficam indisponíveis. Faz parte do modo de circulação do próprio digital.

– *dimensão e heterogeneidade do arquivo*: lidamos com uma infinidade de textos na internet. Quantidade. Paráfrase.

– *autoria*: muitas vezes nos deparamos com materiais que não têm um “nome de autor” ou uma chancela institucional e temos que descartá-los pela demanda de “legitimidade” do arquivo.

– *leitura dispersiva*: a leitura se desloca do fio temporal linear passando a predominar a ordem espacial, na qual se impõe a visualidade (DIAS, 2013, p. 55). Para compreender o arquivo na internet, precisamos compreender a dispersão constitutiva desse arquivo, que impõe um ritmo ao trabalho de leitura.

Apesar de todas essas questões que tocam o problema do método, e que nos levam a colocar questões a propósito do arquivo digital [como lidar com essa dispersão de textos, verbais, não verbais, chancelados ou não?], é preciso considerar que a relação com o arquivo já é determinada *a priori* por uma questão de pesquisa, o que já dá a ele uma configuração na direção da constituição do *corpus*.

Lembrando que o *corpus* resulta de um trabalho de leitura de arquivo, cujo nó central é a relação entre língua e exterioridade, uma remetendo ao jogo, ao equívoco, e a outra, aos efeitos linguísticos materiais na história (PÊCHEUX, 2010, p. 58).

É preciso atentar, ainda, para o fato de que na discursividade da rede, nesse modo particular produção dos discursos, em que ler o arquivo é parte do próprio momento de sua circulação, a textualização é determinada pelo processo de atualização dos sentidos no eixo horizontal. Não se trata da atualização da memória discursiva pela formulação num intradiscorso, mas da atualização de dados pela circulação.

“Isso circula”, como adquirimos o hábito de dizer, fazendo dessa circulação a imagem positiva de nossa modernidade discursiva liberada, ou ao contrário, a falsa moeda das línguas de vento: os turbilhões esfumados do “não importa o que” destinados a chamar a atenção, desviando-a dos “problemas reais”. Não seria tempo de destituir essa imagem duplamente satisfatória da circulação, assumindo o fato de que as circulações discursivas não são jamais “não importa o que”?⁵ (PÊCHEUX, 1981, p. 18)

⁵ Tradução livre de: “‘Ça circule’, comme on a pris l’habitude de dire, en faisant de cette circulation l’image positive de notre modernité discursive libérée, ou au contraire la fausse monnaie de langues de vent; les tourbillons fumeux du ‘n’importe quoi’ destinés à occuper l’attention, en la détournant des ‘problèmes réels.’”

O fato de que “algo circula”, e pelo fato mesmo de sua circulação ser parte de uma massa quantitativa de dados armazenados numa memória metálica (horizontal), constitui o modo das relações entre sujeitos e sentidos. Mas como afirma Pêcheux na citação acima: as circulações discursivas não são jamais “não importa o que” por causa, justamente, da sua materialidade.

Sendo assim, é preciso construir dispositivos de arquivo específicos atentando para as condições de produção do digital e, a partir desses dispositivos, reunir o *corpus*. Ou seja, estamos tratando de todo um trabalho complexo entre a teoria, o método, os procedimentos e o objeto da análise de discurso – o discurso –⁶ que não podem ser negligenciados em prol de um “novo” objeto de análise.

Retorno ao arquivo

Em 1998, como parte da pesquisa que desenvolvia no Mestrado em Letras, na Universidade Federal de Santa Maria (UFSM), criei uma página na internet, que foi hospedada no *site* daquela universidade. Sob orientação da professora Amanda Scherer, a pesquisa tinha como tema os modos de enunciação de si do sujeito homossexual soropositivo. Assim, a criação da página na internet tinha como objetivo criar um espaço de enunciação de si para que sujeitos homossexuais soropositivos falassem do viver com o HIV.

Naquele momento de emergência da internet, e dado todo o estigma em torno da Aids, pensei ser a web a melhor forma de fazer um sujeito, anonimamente, falar de si nessas condições de produção. Criei, então, a página. Era um discurso nativo, como refere Paveau (2014), da internet. Na época, ainda em linguagem HTML. Mas o projeto não deu certo nesse formato. A razão, certamente, deve-se ao fato de que em 1998 não havia a difusão da internet como hoje e nem as mesmas condições de produção da era social da internet. Para termos uma ideia, o Brasil, no final dos anos 1990, tinha pouco mais de um milhão de usuários. Hoje, são mais de 60 milhões.



Figura 1. Página inicial do projeto “Aids e homossexualidade: uma autobiografia virtual (1998)

⁶ Sobre essa questão e a crítica à indistinção, diluição e apagamento das filiações em análise de discurso, ver Orlandi (2012).

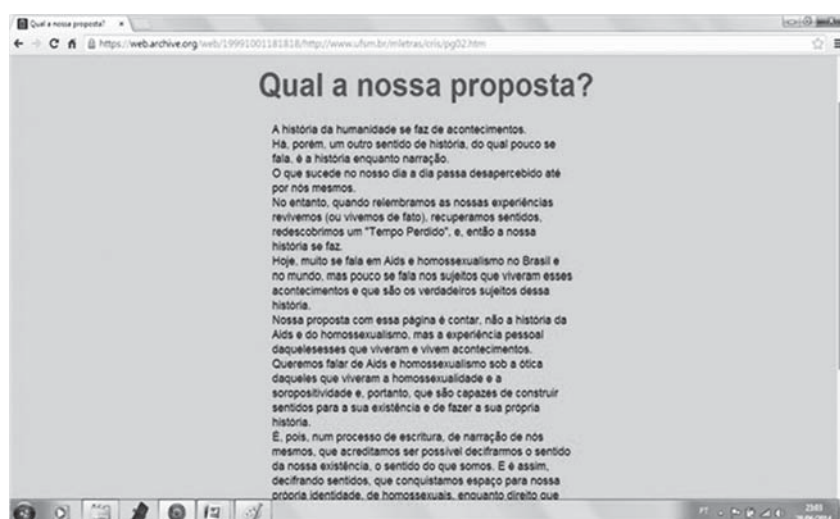


Figura 2. Página segunda do projeto “Aids e homossexualidade: uma autobiografia virtual (1998).

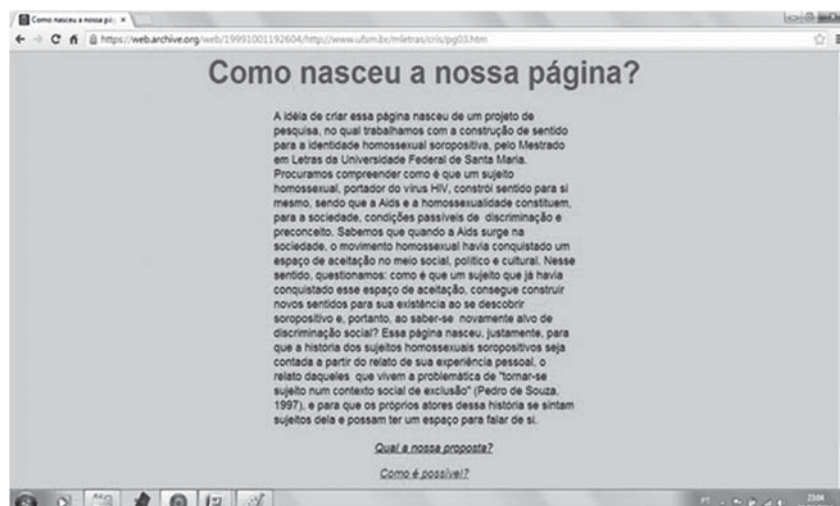


Figura 3. Página terceira do projeto “Aids e homossexualidade: uma autobiografia virtual (1998).

O fato é que hoje, 15 anos depois da criação dessa página, pude fazer um “retorno ao arquivo”, pelo *Internet Archive* –⁷ uma organização sem fins lucrativos dedicada a manter uma espécie de “arquivo da internet”.

O que o *Internet Archive* faz é o que Pêcheux (1981, p. 16) chamou de “leitura-trituração”, que consiste em “recortar, extrair, deslocar, aproximar”. Operações nas quais se constitui um dispositivo muito particular de leitura, diz o autor, produzindo uma “religião do sentido”. Um “espaço da repetição onde tudo é sempre e em toda parte, aproximadamente variável”⁸ (PÊCHEUX, 1981, p. 17).

Eis o que Pêcheux, já em 1981, chamava de um vazio político para o qual a análise de discurso precisa atentar. Não é diferente, hoje, sobretudo, com as facilidades da

⁷ Disponível em: <<https://archive.org/>>. Acesso em: 20 maio 2014.

⁸ Tradução livre de: “un espace de la répétition où tout est toujours et partout comme ici aux variations près”.

internet, em que a “religião do sentido” parece imperar, tanto no que diz respeito a um dispositivo de leitura que se vale das operações de extração, aproximação, recorte, etc., apagando as condições de produção e as relações significantes, quanto no que diz respeito à própria construção de um *corpus* de trabalho de pesquisa sem embate com o arquivo, já que tudo parece “dado”, exposto, um já-lá arquivístico. Assim, é preciso, enquanto analistas de discurso, não descuidar do método, já que, como adverte Orlandi (2012, p. 38) “é preciso, para se fazer ciência do discurso, estabelecer uma relação de consistência entre a teoria, o método, os procedimentos, e o objeto”. E isso consiste num “real trabalho de arquivos” que nada tem a ver com “acúmulo de textos à disposição”, mas com um trabalho de leitura. Nada tem a ver com a reprodução de sentidos, mas com produção de acontecimentos (PÊCHEUX, 1981, p. 17), passíveis de “deslocar as linhas de clivagem” e “inquieta as posições estabelecidas”.

Desse modo, ao “retornar ao arquivo”, por meio da página que criei, disponibilizada pelo *Internet Archive*, o que está em questão nessa retomada?

Esse retorno ao arquivo consiste em considerar a dispersão mesma do arquivo recorrendo ao “trajeto temático”, que “corresponde a um procedimento de compreensão”, fundamentado num “vaivém de atos languageiros” (GUILHAUMOU; MALDIDIER, 1994, p. 165), numa diversidade de textos e de sujeitos. O que está em jogo aqui é a materialidade da internet, hoje, na retomada de minha pergunta de pesquisa na conjuntura atual, que se divide e se desloca na relação com uma memória metálica, cujos efeitos na constituição dos sujeitos e dos sentidos são da ordem do discurso.

Busquei então, levando em conta essa ordem do discurso contemporâneo que, para mim, tem como matéria-prima o falar de si, alguns *blogs* de sujeitos portadores do vírus HIV, dos quais trago aqui apenas dois e destaco, para a análise, o perfil do blogueiro e a descrição do *blog*:

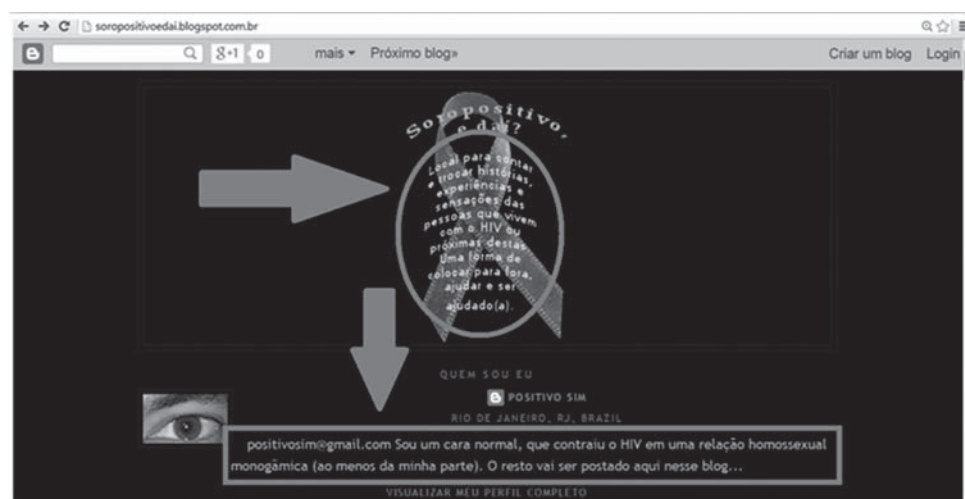


Figura 4. *Blog Soropositivo e daí.*

- (01) Sou um cara normal, que contraiu o HIV em uma relação homossexual monogâmica (ao menos da minha parte). O resto vai ser postado aqui nesse blog...
- (02) Local para trocar e contar histórias, experiências e sensações das pessoas que vivem com o HIV ou próximas destas. Uma forma de colocar para fora, ajudar e ser ajudado(a).



Figura 5. Blog Eu e o hiv.

- (03) BRASIL, Sudeste, BAURU, Homem, English, Portuguese, Cinema e vídeo, Informática e Internet, Vídeos Games. Idade: 39 anos. Profissão: Tradutor. Hobby: Jogar vídeo game.
- (04) Sou apaixonado pela vida, e independente de ser ou não soropositivo desde 95, eu me considero uma pessoa feliz!

Não vou aqui proceder à análise propriamente do falar de si, nas condições de produção contemporânea da internet. Meu objetivo é refletir sobre a constituição do arquivo na história e sobre o gesto de leitura de arquivo na internet, que deve levar em conta a materialidade digital. No recorte que apresento, o que é interessante observar é que as condições de produção da internet em 1998, seja técnica ou socialmente, não havia construído na história o sentido das relações digitais, que se produz ao longo da constituição da internet, com os *blogs*, conversas instantâneas, *chats*, redes sociais.

Sobre isso, diria que esse movimento de retorno ao arquivo demonstra que a construção do arquivo não é linear e nem pontual, mas deve ter como fio condutor uma questão, que tampouco é fixa, mas, ao contrário, assume diferentes formas materiais. Gostaria de mostrar, ainda, que a história não é linear e que os sentidos que nela se produzem não cabem na transparência da história da internet, mas derivam, sim, de um retorno ao arquivo, que, como vemos, não é um retorno ao mesmo, mas um retorno que se dá pela polissemia – *deslocamento, ruptura de processos de significação* (ORLANDI, 1999).

É nesse ponto que o acontecimento discursivo pode ser apreendido na consistência dos enunciados dispersos na rede. Eis o que Pêcheux (1981, p. 17) advertiu sobre “livrar a análise de discurso dos sulcos da reprodução do sentido e engajá-la na produção de acontecimentos”.⁹

⁹ Tradução livre de: “dégager l’analyse des discours des ornières de la reproduction du sens, et l’engager dans la *production des événements*”.

Conclusão

Pensar o arquivo, a leitura e a constituição do *corpus* da perspectiva da análise de discurso é de suma importância, pois, com a internet, o discurso digital, pensa-se que tudo é novo, incorrendo naquilo que Orlandi (2012, p. 29) chamou de “narrativa das filiações”:

Antes mesmo de aprenderem a fazer uma análise, há muitos estudantes e jovens professores que começam por “criar” teoria ou a fazer a história da análise de discurso, sem um real trabalho de arquivos. Colocam a mochila nas costas e saem na busca de novas possibilidades (o importante é chamar de “novo”).

Ora, é fato que o funcionamento dos discursos, com a internet e o próprio trabalho com o arquivo, precisa levar em conta as condições de produção da internet e, como venho dizendo, a discursividade da rede de sentidos, que não escapa à injunção do digital e dos modos de existência dos sujeitos e de produção dos sentidos na sociedade digital. Para isso, porém, é preciso um real trabalho de leitura de arquivo, ou seja, é preciso compreender a relação língua e discursividade. O resto é, como diz Orlandi (2012), fabulação.

REFERÊNCIAS

- DIAS, Cristiane. Linguagem e tecnologia: uma relação de sentidos. In: PETRI, Verli; DIAS, Cristiane (Org.) *Análise de Discurso em perspectiva: teoria, método e análise*. Santa Maria: Editora da UFSM, 2013. p. 49-62.
- GUILHAUMOU, Jacques; MALDIDIER, Denise. Effets de l’archive. In: GUILHAUMOU, Jacques et al. *Discours et archive: expérimentations en analyse du discours*. Liège: Mardaga, 1994. p. 91-111.
- ORLANDI, Eni. Segmentar ou recortar? *Linguística: questões e controvérsias*. Série Estudos 10. Curso de Letras do Centro de Ciências Humanas e Letras das Faculdades Integradas de Uberaba, 1984.
- _____. *Análise de discurso: princípios e procedimentos*. Campinas: Pontes, 1999.
- _____. Apagamento do político na ciência: notas à história da análise de discurso-fragmentação, diluição, indistinção de sentidos e revisionismo. In: _____. *Discurso em análise: sujeito, sentido, ideologia*. Pontes: Campinas, 2012. 239 p.
- PAVEAU, Marie-Anne. Les énoncés natifs du web: analyse du discours des réseaux sociaux numériques (Twitter, Facebook, Pinterest). Campinas: Unicamp, 2014. Disponível em: <<http://www.labeurb.unicamp.br/anexos/MAP-Conf.pdf>>. Acesso em: 22 nov. 2014.
- PÊCHEUX, Michel. Ouverture du colloque. In: CONEIN, Bernard et al. (Org.) *Matérialités discursives*. Lille: Presses Universitaires de Lille, 1981. p. 15-18.
- _____. *O discurso: estrutura ou acontecimento*. Tradução de Eni Orlandi. 5. ed. Campinas: Pontes, 2008.
- _____. Ler o arquivo hoje. In: ORLANDI, Eni (Org.) *Gestos de leitura*. 3. ed. Campinas: Editora da Unicamp, 2010.