

Dize-me qual é teu corpus, eu te direi qual é a tua problemática

Revista Diadorim / Revista de Estudos Linguísticos e Literários do Programa de Pós-Graduação em Letras Vernáculas da Universidade Federal do Rio de Janeiro. Volume 10, Dezembro 2011.

Os problemas colocados pela noção de corpus são relativamente bem conhecidos por terem sido objeto de longas discussões no campo das ciências da linguagem [1], mas que nem sempre resultaram num consenso que permitisse recorrer a uma definição consolidada à qual se pudesse referir cada vez que se tivesse de justificar um corpus de análise. Talvez isso seja o sintoma de que o corpus não existe em si, mas depende – como mostra o artigo “corpus” do Dicionário de Análise do Discurso – do posicionamento teórico a partir do qual ele é considerado. Tratando-se da linguagem, são bem conhecidos os problemas que a constituição de um corpus implica, dentre os quais citaremos apenas os principais :

1. o problema que concerne à coleta de dados, coleta que depende da escolha da materialidade linguística (produções linguísticas orais, produções linguísticas escritas), da escolha do suporte que veicula tais produções em relação com uma situação de comunicação (para o escrito : cartas, relatórios, jornais, panfletos, circulares, cartazes etc. ; para o oral : rádio, televisão, reuniões diversas, comícios, conversas quotidianas etc.). Todos eles têm incidências sobre a maneira de coletar dados : exploração do terreno, procedimentos de gravação livres ou controlados, com conhecimento ou não dos atores da fala, etc. ;
2. o problema que concerne à importância do material coletado e de seu valor de representatividade ; o corpus pode ser considerado exaustivo e fechado, ou parcial e aberto, e, conseqüentemente, pode ser considerado como um objeto em si ou uma simples ferramenta (MAYAFFRE, 2005). Sabe-se que a hipótese da exaustividade – velho sonho da atitude positivista – não se sustenta mais, e isso, apesar do desenvolvimento da chamada linguística de corpus iniciada no mundo anglo-britânico e que foi seguida na França por alguns autores com uma certa prudência, pois a linguagem está em perpétuo processo de produção. É possível, entretanto, que o fechamento do corpus seja reivindicado, a título experimental, por alguns analistas. Se o corpus é considerado como parcial, coloca-se então o problema de seu valor enquanto amostra, e da possibilidade da constituição de subcorpus. Quanto a considerar se o corpus é um objeto em si ou um instrumento, isso remete à questão do contexto que analisaremos mais adiante ;
3. o problema que concerne, no interior do material linguístico, às categorias objeto da análise : gramaticais (conectores, pronomes, verbos, etc.), lexicais (por campos ou de maneira aleatória), sintáticas (segundo diversos tipos de construção) ; mas também as variáveis externas à produção dos atos de linguagem, como os tipos de locutores, os dispositivos de comunicação, tanto quanto as variáveis que dizem respeito ao tempo (a historicidade) e ao espaço (as culturas).
4. o problema, enfim, que concerne à ferramenta de tratamento dos dados : seleções manuais, tratamento informatizado com o uso de softwares ad hoc, constituição de amostras a partir de bases de dados (ver abaixo).

Em todos esses casos, colocamos a questão da pertinência dessas escolhas em relação com os pressupostos teóricos, e a da eventual circularidade que pode se instaurar entre estes e o corpus.

Gostaríamos de mostrar, neste artigo, de que maneira a construção de um corpus, na análise de discurso, depende de um posicionamento teórico ligado a um objetivo de análise, o que chamamos de “problemática”. Abordaremos três tipos de corpus segundo se refiram a uma problemática dita cognitiva, comunicativa ou representacional.

Posteriormente, mostraremos como um corpus de textos se constrói num jogo de desconstrução e de reconstrução segundo critérios de abertura ou fechamento (intertexto e hipertexto), de contraste (interno e externo) e de pertinência quantitativa e/ou qualitativa. Terminaremos pela menção a um exemplo de análise praticado pelo CAD a propósito do discurso de informação da televisão, a fim de mostrar como podemos tirar partido de algumas dessas opções, e como elas podem se combinar.

Antes, porém, é necessário circunscrever a especificidade da análise do discurso no âmbito das ciências da linguagem, pois não podemos tratar da questão dos corpora de discurso sem indagar a respeito de certas noções utilizadas com frequência nesse domínio.

Sobre algumas distinções indispensáveis para formar corpus de discurso

Nas ciências humanas e sociais, algumas disciplinas privilegiam estudos de campo com procedimentos que seguem um método descritivo-empírico (sociologia, antropologia), outras privilegiam procedimentos experimentais (psicologia social), outras, análises a partir de arquivos ou a partir de corpora organizados com procedimentos mais ou menos sistemáticos de coleta e tratamento do material semiológico reunido (história, ciências da linguagem).

As ciências da linguagem fazem parte, então, das disciplinas de corpus : compilação de dados linguísticos (sob a forma de textos escritos ou orais, de documentos diversos, de observações empíricas selecionadas ou de sondagens provocadas) que são constituídos em objeto de análise. Assim sendo, colocamos a questão de saber qual é a natureza desses dados. Não podemos, então, tratar da questão do corpus se não estivermos esclarecidos (isto é, se não tomarmos posição) a respeito de algumas noções que estão no cerne das ciências da linguagem, das quais abordaremos aqui algumas distinções que me parecem as mais fundamentais : língua/discurso, texto/discurso, texto/contexto.

Língua/discurso

Questão superada, dirão alguns. No entanto, apesar de o próprio Saussure declarar que a língua só é criada a partir do discurso, muitas definições da noção de discurso, que foram propostas desde os anos 1970, obscurecem sua oposição à língua. Para demonstrar isso, basta remeter a Maingueneau que, em 1976 e depois em 1996, fornece um repertório de definições. Algumas fazem do discurso um prolongamento da língua, opondo a unidade da frase a uma unidade "transfrástica" (Harris 1969), outras integram categorias de discurso à língua, outras ainda opõem discurso a enunciado, relacionando essa noção às "condições de produção" de um texto ; outras, enfim, trazem para o termo "discurso" um uso que Maingueneau qualifica de "paralinguístico", na medida em que se integra a uma reflexão que participa de várias disciplinas, tais como a história, a filosofia, a semiologia, e mesmo a psicanálise.

Aqui não é o lugar de discutir essas definições. Trata-se somente de marcar a diferença entre língua e discurso no que ela concerne ao ponto de vista da análise e da constituição de seu objeto. Essas duas noções assinalam dois lugares de estruturação da linguagem :

1. a língua, como lugar de consubstanciação entre formas e sentido organizando-se em sistemas, isto é, em redes de relações entre unidades mínimas segundo regras de combinações sintagmáticas e paradigmáticas, sistemas dos quais se poderá dizer que remetem a categorias de pensamento que tomam posição sobre visões do mundo. As unidades que são levadas em conta são unidades fonológicas, morfológicas, semânticas, de ordem gramatical ou lexical, e os corpora são constituídos por um conjunto de co-ocorrências por semelhanças de forma ou de sentido.
2. o discurso, como lugar, ao mesmo tempo, de estruturação dos usos em função das condições de produção nas quais esses usos se manifestam, relacionados a comportamentos languageiros dos sujeitos falantes, e categorizações de sentido relacionadas a sistemas de conhecimento e de crença aos quais aderem os indivíduos ou grupos sociais. Aqui, não se trata mais de uma combinação de unidades em nível transfrástico, mas, de uma dimensão para além da frase, uma vez que o sentido de discurso resulta de uma multiplicidade de fatores de ordem contextual (ver adiante) que fazem com que o sentido não seja observável em nenhuma das unidades que compõem uma produção discursiva qualquer, e só pode ser inferida pelo jogo de combinações de todos esses fatores. Para retomar uma noção de Roland Barthes, o sentido se liga às formas por "difração". Um corpus de discurso só pode ser constituído por um conjunto de produções languageiras em situação de uso.

Por conseguinte, dá-se a constituição de duas linguísticas, uma que se chamará de "linguística da língua", orientada para a descrição dos sistemas intrínsecos a cada língua, com um corpus mais ou menos fechado, e que pressupõe a existência de um sujeito operador de categorias "cognitivo-linguísticas", outra que se chamará de "linguística do discurso", orientada para a descrição dos usos e das significações sociais, com um corpus, por definição aberto, e que pressupõe um sujeito operador de categorias "sócio-discursivas" e portador de um imaginário social.

Vejamos um exemplo para ilustrar essa diferença. Um enunciado como "Eu tenho trinta anos" poderá ser o objeto de uma análise linguística de língua quando, inserido num corpus de enunciados similares conforme o esquema < Pr. x Vb. x quant. x Sb. >, o analista determinará o actante (aquele que fala) ao qual é atribuída (ter) certa propriedade (anos), a qual é quantificada (trinta), e que constitui um ato de enunciação que diz que essa asserção deve ser atribuída ao próprio sujeito falante (Eu) numa modalização elocutiva afirmativa ; além disso, observando um certo paradigma lexical, diremos que "anos" designa um certo segmento do tempo que remete à idade da vida. Mas globalmente, "Eu tenho trinta anos" não significará aqui nada mais do que /eu tenho trinta anos/.

Se considerarmos agora esse enunciado em seu contexto de produção, a saber, um desportista respondendo "Eu tenho trinta anos" à reflexão de um amigo que se espanta ao saber que ele deixará de competir, então esse enunciado significa, entre outras coisas : /eu estou muito velho/. Mas se fosse dito por uma pessoa que, acabando de ser demitida de uma empresa, responde a um amigo que lhe sugere que é talvez porque já tenha passado da idade, então o "Eu tenho trinta anos" significa : /mas eu ainda sou jovem !/ Esses sentidos dependem evidentemente da sequência das palavras contidas no enunciado, mas foram obtidos por uma operação de inferência a partir do conhecimento que o interlocutor pode ter da identidade daquele que fala (desportista/empregado) e do universo de sentido ao qual remete cada um deles (esporte/empresa), elementos esses incluídos num corpus memorial de discurso.

Resta que o sentido de discurso, se ele é inferido, o é a partir de formas cuja combinação constitui um texto. Trata-se então de saber se texto e discurso são uma só e mesma coisa.

Texto/discurso

Tendo em vista a quantidade do que já foi escrito sobre essas noções, e mesmo sobre uma distinção que nem sempre é efetuada, delas nos ocuparemos de maneira breve. Vamos somente fazer uma constatação e tomar posição, posto que isso é necessário para determinar um corpus de discurso.

A constatação : raramente as noções de texto e discurso se opõem de maneira clara, a tal ponto que obras que apresentam abordagens da "análise textual" empregam com frequência a expressão "análise de discurso" e inversamente. Consciente da questão, Jean-Michel Adam propõe "distinguir texto e discurso como as duas faces complementares de um objeto comum assumido pela linguística textual", ponderando que esta última privilegia a organização do contexto segundo critérios de coesão, enquanto a análise do discurso se ocupa mais do contexto segundo critérios de coerência. Essa distinção mostra bem que o objeto texto se define pela organização de sua configuração em relação com o que o cerca e que o discurso diz respeito à organização semântica, embora o texto também seja portador de sentidos e o discurso também seja acessado através de formas.

Essa distinção, entretanto, não é clara o bastante para diferenciar corpus de textos e corpus de discurso. É mais : se adotarmos a série de noções propostas por G. Genette – paratexto, metatexto, epitexto, intertexto e hipertexto –, às quais podemos acrescentar a noção de interdiscurso introduzida pela escola francesa de análise do discurso dos anos setenta/ oitenta, não vemos com clareza onde se encontra a fronteira entre essas duas noções.

A tomada de posição : não temos um ponto de vista muito diferente daquele de Jean-Michel Adam, exceto que consideramos que a relação de complementaridade entre análise de texto e análise de discurso não é simétrica. Não são as duas faces de uma mesma moeda, porque há discursos diversos num mesmo texto e um mesmo discurso em vários textos. O discurso não é o texto, mas ele é carregado por textos. O discurso é um percurso de significância que se acha inscrito num texto, e que depende de suas condições de produção e dos locutores que o produzem e o interpretam. Um mesmo texto é então portador de diversos discursos e um mesmo discurso pode impregnar textos diferentes. Há discurso atravessando textos diferentes, e um mesmo texto pode ser portador de discursos diferentes.

Para compreender o slogan publicitário "Obernai, a primeira grande cerveja com um terço de calorias a menos", somos levados a mobilizar uma série de discursos que não estão ditos explicitamente nesse slogan, mas que aí estão implícitos, a saber : /eu devo ter conhecimentos sobre dieta/, /essa cerveja não engorda/, /a esbeltez é valorizada em minha sociedade/, /as mulheres podem beber essa cerveja, tradicionalmente reservada aos homens/. Isso não impede que os discursos sobre dieta, esbeltez e seu valor cultural possam estar configurados em outros tipos de texto dos mais eruditos aos mais vulgares. Como o discurso necessita de configuração textual para significar, isso quer dizer que essa significância, num dado momento, foi texto. Produz-se então um fenômeno estranho de vaivém entre diferentes textos que fazem eco uns aos outros, que acaba construindo uma significância abstrata que se encontra em diferentes textos sem estar num deles unicamente. Para dar um exemplo simples, o discurso do pecado original e o imaginário de culpabilidade e de redenção de que é portador, construíram-se no Ocidente pelo "ecoar" entre diversos textos (Bíblia, Evangelhos, obras de teólogos, comentários de filósofos, pregações da Igreja, chegando até os textos publicitários, as revistas populares, os filmes, etc.). Assim são depositados discursos na memória coletiva.

Texto/contexto

Ao considerar o discurso como um lugar de significância abstrato configurando-se diversamente em textos, retomamos as noções de intertextualidade de Genette e de dialogismo de Bakhtin. Partindo desta última noção que suscitou diversas interpretações, diremos que esse jogo de ecos pode operar-se entre textos mas também entre discursos. No primeiro caso, trata-se, no sentido

estrito do termo, de uma intertextualidade que se manifesta em diversos tipos de citação – o que Jacqueline Authier-Revuz chama de “heterogeneidade mostrada” : há texto dentro do texto, mesmo parcial, mesmo deformado. No segundo caso, trata-se de uma interdiscursividade, “um jogo de remissões entre discursos que tiveram um suporte textual mas cuja configuração não foi memorizada”. Isso nos explica que um corpus só possa ser formado por textos e não por discursos, a menos que se decida que os textos reunidos o foram em nome de um certo discurso. Mas isso será visto mais adiante.

Chegamos então à noção de contexto. Desde os primeiros trabalhos em linguística da língua, a noção de contexto estendeu-se consideravelmente com denominações variáveis. Em resumo, fala-se de : contexto estritamente linguístico, quando se trata de se referir à vizinhança verbal que se encontra antes e/ou depois de uma determinada unidade ; de contexto textual, quando este é constituído de textos produzidos por uma mesma fonte (por exemplo, as diversas obras de um mesmo autor) ; de contexto paratextual (Genette), quando é constituído de textos ou fragmentos de textos que se acham em copresença num mesmo espaço material de escrita (os títulos, subtítulos, lides, legendas, de uma página de jornal) ; de contexto metatextual (Genette), quando é constituído de textos que se comentam uns aos outros ; de contexto hipertextual (Genette), quando é constituído de textos que se citam, remetem uns aos outros, retomam e transformam uns aos outros, como os pastiches e as paródias ; enfim, de contexto intertextual e interdiscursivo (cujo sentido acabamos de definir), quando são convocados textos e discursos para justificar as interpretações do sujeito interpretante (ou analisante).

Assim sendo, nesse alargamento progressivo da noção de contexto, aparece uma tomada de consciência progressiva, não somente da relação entre texto e toda a vizinhança textual que está relacionada a ele, mas também entre o texto e um “fora do texto” (às vezes chamado de cotexto), isto é, os dados presentes nas condições de produção do ato de linguagem.

O que gostaríamos de destacar aqui é um aspecto raramente levado em consideração pelas diferentes correntes de análise do discurso, e que nos permite introduzir a noção de contexto situacional. Trata-se do aspecto que concerne, nas condições de produção, não somente aos pressupostos de posicionamento interdiscursivo (pressupostos ideológicos, segundo alguns autores), mas também ao condicionamento da própria situação de comunicação : a natureza da identidade dos parceiros do ato de linguagem, a finalidade da situação, os dispositivos e as circunstâncias materiais desta última. Esse conjunto constitui o que chamamos de contrato de comunicação, o qual restringe o sujeito falante, dando-lhe as instruções discursivas que deverá seguir para proceder a seu ato de enunciação.

É certo que os etnógrafos e etnólogos da comunicação interacional nos habituaram a levar essas variáveis em consideração para analisar as conversações, mas tais variáveis não são suficientemente consideradas na análise do discurso. É que a análise do discurso tem trabalhado, de preferência, com a hipótese de que a força da linguagem se encontra mais no que dizem as palavras do que naquilo que elas transmitem ao serem ditas.

As pesquisas se concentraram, então, no que a linguagem diz, sem colocar a questão do poder da linguagem em exercer influência. Melhor dizendo, para retomar uma observação de Roland Barthes, pensa-se sempre que o signo “significa” e se esquece de que ele “significa para”. Ora, o fenômeno de significância resulta desses dois processos : a linguagem significa ao mesmo tempo em que transmite ; é no próprio ato de transmissão que ela significa, com o efeito que produz sobre o outro, isto é, sua interpretação. Não colocar a questão dos efeitos é amputar o ato de linguagem de sua significação fenomênica, a saber que este ato é o resultado de uma construção.

Só se pode captar o sentido de uma expressão como “preferência nacional” se for considerada em seu contexto situacional, o qual nos dirá quem a emprega (Jean-Marie Le Pen), destinada a quem (cidadãos-eleitores), em qual dispositivo comunicacional (campanha eleitoral), com qual finalidade (seduzir), e, evidentemente, relacionando-o a uma certa tradição histórica francesa (interdiscurso).

O problema que se coloca então é o da articulação entre, de um lado, aquilo a que remetem os discursos do ponto de vista dos valores, e do outro lado, os efeitos que os discursos são capazes de produzir. Por exemplo, o poder de influência das palavras do discurso político não está somente na força do projeto de ideal social de que são portadores, mas também, e cada vez mais, em seu eco nos sistemas de expectativa dos cidadãos (efeito de espelho), os quais são mais sensíveis ao carisma das personalidades políticas, à emoção que estas suscitam, do que às ideias. O contexto situacional deve então ser levado em conta tanto quanto os demais para formar corpus. Evidentemente, nenhum corpus pode levar em conta a totalidade desses contextos, mas é considerando essa multiplicidade contextual que se poderá construir corpus de maneira racional.

Especificidades do corpus em análise do discurso

A considerar estas diferentes distinções – o discurso como significância/o texto como configuração e o contexto sob seus três aspectos (paratextual, interdiscursivo e situacional) –, justifica-se o fato de que um corpus de discurso seja sempre uma construção resultante de diversos tipos de agrupamentos : corpus segundo o paratexto (de palavras, de enunciados, de modos de enunciação), corpus segundo o interdiscurso (saberes de conhecimento, saberes de crença), corpus segundo a situação (locutores, finalidade e dispositivo). Esses agrupamentos se fazem em função da problemática de análise e do tipo de contraste escolhido.

Corpus e problemática de análise

Não utilizarei o termo “problemática” com o sentido que este apresenta no uso corrente, como na expressão “Como é problemática !”, para significar que uma questão é difícil de explicar, ou sem solução, ou que uma afirmação é incerta, duvidosa ou mesmo suspeita. Uma problemática (emprego substantivo) é um conjunto coerente de proposições hipotéticas (ou de postulados) que, no interior de um campo de estudo, determinam ao mesmo tempo um objeto, um ponto de vista de análise e um questionamento por oposição a outros questionamentos possíveis. Não se confundirão, então, problemática e teoria, que se diferenciam pelo fato de que esta última tem um corpo de conceitos e de categorias mais ou menos padronizado, acompanhado de um quadro metodológico.

Observando os estudos que se relacionam de perto ou de longe com o discurso, e interrogando-os segundo o tipo de objeto que analisam e a concepção que adotam a respeito do sujeito que o organiza, podemos distinguir três grandes problemáticas que chamaremos de cognitiva, comunicativa e representativa, cada uma determinando um tipo de corpus.

Uma problemática dita cognitiva e categorizante

O objeto de estudo é considerado como um conjunto de mecanismos discursivos. Trata-se, então, ou de captar sua existência e seu modo de combinação no interior de uma produção discursiva qualquer (texto ou enunciados aleatórios), ou de descrever como é que se dá seu funcionamento, e particularmente seu modo de produção. No primeiro caso, estuda-se, por exemplo, os meios de coerência e coesão de um texto, ou algumas palavras do discurso (conectores, modalidades). O segundo caso é encontrado nos trabalhos de alguns psico-cognitivistas da linguagem que procuram descrever percursos de organização discursiva “down” ou “up” de realização dos esquemas, scripts e roteiros. Nos dois casos, encontram-se categorias instrumentais de análise que servem para discernir ou produzir configurações discursivas.

O sujeito ligado à determinação desse objeto será chamado cognitivo, na medida em que só é levada em consideração sua aptidão em produzir ou discernir operações de articulação discursiva : relações anafóricas ou catafóricas, conexões coordenadas ou subordinadas, relações de coerência e de coesão, de progressão e de não contradição, etc., sem levar em conta sua natureza social.

Numa tal problemática, o corpus não necessita ser estabelecido em função de uma situação de comunicação particular. Desse ponto de vista, ele pode ser considerado aleatório, e somente limitado pelos contextos linguísticos nos quais aparecem esses mecanismos. Constrói-se um corpus de co-ocorrências por categorias de operação linguístico-discursiva.

Pertencem a essa problemática trabalhos na linha das teorias da pressuposição e do sentido das palavras do discurso, dos topoi e da orientação argumentativa e da pragmática (os conectores, os em-breadores⁴ e os atos de fala), assim como os trabalhos em gramática de texto que tratam dos problemas de coerência e coesão textuais.

Uma problemática dita comunicativa e descritiva

O objeto de estudo é aqui empírico, ou mais exatamente é determinado a partir da observação (mais ou menos ingênua) das manifestações do mundo fenomênico. Por exemplo, podemos observar que os comportamentos dos indivíduos em seus atos de troca social têm uma certa recorrência, o que permite formular a hipótese de que são submetidos a condições de realização e que obedecem a regras ; e é isso que autoriza a constituir-los em objeto de análise.

Assim sendo, o objeto de estudo oriundo dessa empiria pode ser estruturado em “tipos ideais” (no sentido durkheimiano) de comunicação, a partir dos quais são descritas as características discursivas que se ligam a cada um deles (identidade dos locutores, finalidade, dispositivo). Tal procedimento resulta, entre outras coisas, no estabelecimento de diversas tipologias ditas de discurso, de textos, de gêneros, ou de situações comunicativas, mas que supõem, todas, uma teorização desses tipos ideais de comunicação. Essa problemática pode ser chamada de descritiva pelo fato de que visa a descrever tipos de situação comunicativa.

O sujeito, aqui, está ligado à empiria das trocas comunicativas, mas também está construído e teorizado em função da maneira pela qual se constroem e teorizam essas trocas. O sujeito é, pois,

um sujeito de comunicação que se define por sua identidade psicológica e social, por um comportamento que é finalizado ao mesmo tempo pelas restrições a que obedece ao se inserir na troca (desse ponto de vista é indeterminado, sendo um sujeito da norma contratual) e por suas próprias intenções com relação ao outro (desse ponto de vista é um "eu", um sujeito que se torna um indivíduo).

O corpus se ressentir dessa visão empírica do objeto de estudo. Geralmente é constituído de textos (produções linguageiras em situação) que são agrupados em função de seu pertencimento a tal ou qual tipo de situação. Por exemplo, textos publicitários, textos jornalísticos, textos de manuais escolares, textos administrativos, textos de programas políticos, e diversos textos conversacionais (trocas telefônicas, pedidos de informações, entrevistas, debates, etc.). Esse tipo de agrupamento permite proceder a comparações em torno de um mesmo contexto paratextual e situacional. Pertencem a essa problemática trabalhos que abordam a descrição dos gêneros do discurso e dos objetos empíricos categorizados em situação de comunicação.

Uma problemática dita representacional e interpretativa

O objeto de estudo dessa problemática é definido através das hipóteses de representações sócio-discursivas que se supõem dominantes num dado momento da história de uma sociedade (são, então, sócio-históricas), e que caracterizam um determinado grupo social. Sob esse aspecto, elas são interpretativas, visto que é necessário formular, de início, uma hipótese sobre o que são os "posicionamentos sociais" em relação com as "práticas discursivas" e os "tipos de sujeitos" que se acham ligados a tais posicionamentos e práticas.

A dificuldade dessa problemática reside justamente no fato de que é preciso formular essas hipóteses e que, para isso, é necessário basear-se em manifestações discursivas extremamente diversas, nem sempre explícitas, e mesmo difusas ou imprecisas, o que é um problema para a constituição do corpus.

O sujeito, nessa perspectiva, também constitui um problema e tem sido frequentemente discutido. Pelo fato de que está na origem das práticas discursivas que constroem as representações, ele pode ser considerado como um sujeito passivo ou ativo. Daí, temos duas posições :

1. uma, radical, que não confere existência a esse sujeito, a não ser a de uma "ilusão", pois ele estaria completamente sobredeterminado pelo que Pêcheux chama de "pré-construído" das "formas discursivas". O sujeito não se pertence, ele é falado por um alheio que será chamado ora de ideologia (Althusser), ora de inconsciente (Authier-Revuz), diluindo-se na consciência do grupo social. Em ambos os casos, o sujeito não é um "eu", mas um "isso" (ideológico ou inconsciente) que fala através do sujeito falante, e não é por acaso que os defensores dessa posição têm feito sempre um paralelismo entre ideologia e inconsciente.
2. a outra posição é menos globalizante. Não nega o jogo da sobredeterminação do sujeito, mas, em vez de fazer do sujeito uma ilusão, faz algo de positivo. Existem, em toda sociedade, discursos de doxa (lugares comuns, estereótipos, saberes compartilhados). Estes circulam e se repartem de maneira não aleatória em diferentes grupos sociais. Assim sendo, todo sujeito é portador de alguns desses discursos, e essa condição revela seu posicionamento social. Podemos então dizer que, de uma certa maneira, esse sujeito é "responsável" por suas representações, o que não quer dizer que seja consciente disso. Na linha dessa posição que qualificaremos de sociológica, acham-se os trabalhos de alguns sociólogos franceses que se inscrevem numa sociologia construtivista, assim como os que trabalham para analisar as representações do espaço público nas mídias.

O corpus varia segundo uma ou outra posição. Na primeira, o corpus é constituído por um conjunto de textos-arquivo que são selecionados por seu valor emblemático de discurso dominante, mais frequentemente ligados a uma instituição social : dicionários, enciclopédias, produções discursivas de grupos constituídos ("cahiers de doléances" [2], partidos políticos, sindicatos) e todas as produções que têm valor de discurso fundador.

Na segunda posição, o corpus é constituído por um conjunto do que chamaremos de signos-sintomas que representam de maneira emblemática sistemas de valores. Esses signos podem ser palavras ("racismo", "imigração", "solidaridade" etc.) ou fórmulas diversas ("purificação étnica") reveladoras de maneiras de dizer ; podem ser também signos icônicos (encenações com imagens) que permitem estudar, por exemplo, as "representações da mulher" nas publicidades ou da "violência" na televisão.

O corpus é aqui de ordem interdiscursiva, mas poderia incluir outros componentes na medida em que é possível aventar a hipótese de que os diferentes elementos de uma situação de comunicação (identidades, finalidade, dispositivos) estão igualmente imersos em valores e ideologias. Por exemplo, se compararmos a maneira pela qual a imprensa organizava suas rubricas e seus gêneros de escrita há 50 anos, com a atual, podemos ver que a maneira de

desempenhar seu papel na simbólica democrática mudou. O que nos remete à problemática precedente.

Pôr em contraste : um jogo de desconstrução/reconstrução

Seguindo a hipótese saussuriana que diz que o sentido nasce da diferença, o sentido de discurso só pode ser captado em operações de contraste, quando se colocam em paralelo diversas sequências discursivas. Assim, os corpora devem ser construídos segundo certas variáveis que permitam comparar tais sequências, quer sejam variáveis externas ou internas.

As variáveis externas possibilitam contrastar conjuntos textuais ou discursivos pertencentes a épocas diferentes (variável temporal) : por exemplo, a imprensa do século XIX comparada à atual. Ou então, esses conjuntos podem ser oriundos de espaços diferentes (variável espacial e cultural) : por exemplo, a publicidade francesa comparada à publicidade brasileira. Ou ainda, podem pertencer a dispositivos situacionais diferentes (variável de gênero) : por exemplo, o discurso político e o discurso publicitário, a fim de saber se as estratégias de persuasão são as mesmas.

Esses corpora em contraste possibilitam trazer dados concretos para resolver as discussões que opõem diversos pontos de vista : o que defende a perenidade dos discursos ("é sempre a mesma coisa através do tempo e do espaço"), em oposição ao que defende a qualquer preço a mudança dos discursos ("nunca é igual de uma época para outra") ; ou ainda o ponto de vista que toma globalmente diversos gêneros discursivos : "tudo isso (discurso político, discurso publicitário), não passa de propaganda", oposto ao que os distingue. Semelhanças e diferenças entre as características de diversos corpora permitem verificar as mudanças ou a perenidade dos discursos, sua generalidade ou sua especificidade quanto ao contexto de produção ou quanto ao gênero.

Mas há também variáveis ditas internas na medida em que estas se situam no interior de um mesmo campo de discurso e correspondem aos componentes situacionais que estruturam um domínio de práticas sociais. O campo político, por exemplo, se compõe de diversos domínios de atividade : os do governo, dos parlamentos, dos partidos, dos Conselhos (constitucional, de Estado, econômico e social) etc. A palavra não circula entre eles do mesmo modo, as normas contratuais não são as mesmas, as estratégias discursivas não se manifestam da mesma maneira, visto que o sentido das palavras varia segundo essas situações. Sem contar o fato de que o discurso político não é o mesmo, se os atores se encontram em situação de conquista do poder (campanhas eleitorais) ou de exercício do poder, a primeira levando a produzir discursos de promessa que são da ordem do desejável, a segunda, discursos de justificativa que são da ordem do possível. Convém, então, construir, no interior de um vasto corpus político, sub-corpus com contextos situacionais diversos.

Há também uma outra maneira de operar contrastes com variáveis internas. Trata-se daquela que consiste em proceder a aproximações entre textos, tratando das mesmas noções. É o trabalho que se faz quando, ao analisar declarações de atores políticos, procura-se compreender o sentido atribuído a palavras como "autoridade", "igualdade", "dever" e "direito". É então necessário recorrer a escritos filosóficos que tratem da república ou da democracia.

Essa variedade de corpus segundo diversas problemáticas às quais eles estão ligados e segundo um jogo de desconstrução/construção resultando em sub-corpus, nos remete a duas questões que são largamente discutidas em análise do discurso : a questão da abertura/fechamento do corpus ; a questão de seu tratamento quantitativo/qualitativo.

Uma questão heurística : a abertura/fechamento do corpus

Resumindo as considerações precedentes, diremos que um corpus depende do tratamento que decidimos dar a ele (sob influência de uma problemática e de um certo jogo de construção em contraste). Poderá ser argumentado que se trata aí de uma circularidade pouco demonstrativa, mas como fazer de outro modo ? Não poderá ser dito que "É finalmente o corpus que faz a teoria", pois seria negar o duplo movimento dedutivo-indutivo da pesquisa em ciências humanas e sociais, mas dir-se-á : "Dize-me qual é o teu corpus, eu te direi qual é tua problemática / dize-me qual é a tua problemática, eu te direi qual é o teu corpus".

O corpus participa assim de uma abordagem heurística, pois propõe um certo caminhar intelectual a partir de hipóteses de base para descobrir "fatos e ideias", isto é, para interpretar. Ora, interpretar é uma operação delicada que só pode ser feita num momento posterior, mesmo quando numa abordagem mais experimental se formulam hipóteses prévias. Interpretar consiste em uma operação de correlação de fatos observados com outros fatos, para formular, por inferência, hipóteses de sentido.

Mas há uma outra atividade interpretativa, a que consiste em relacionar os resultados de uma análise descritiva com os de outras análises : os de outros corpora conexos (confrontação dos

artigos de diferentes jornais para interpretar suas semelhanças e diferenças) ; os de corpora de textos de um mesmo domínio mas de situações diferentes (confrontação de produções jornalísticas de diferentes épocas) ; aqueles, enfim, das análises propostas por outras disciplinas sobre o mesmo domínio discursivo (filosofia, história, sociologia, psicologia social), sobre, por exemplo, o domínio político.

Nas ciências humanas e sociais, a abordagem é dupla : “empírico-descritiva” e “hipotético-dedutiva”. Uma não existe sem a outra : a primeira depende mais de ferramentas metodológicas ; a segunda, de conceitos fundadores e de categorias explicativas. A primeira se desenrola num movimento centrípeto ; a segunda, num movimento centrífugo, o que explica que os corpora que se inscrevem nesses movimentos tendam, ora a se fechar em si mesmos, ora a se abrir.

É por isso que um corpus não passa de um pretexto, no sentido de que ele é um ponto de partida indispensável para uma análise do discurso, devendo os textos ser confrontados permanentemente com outros textos e outros corpora. Isso porque as significações sociais se repartem numa larga combinatória de palavras em relação com seus diferentes contextos. É também por isso que nenhum corpus é definitivamente fechado, constituindo-se seu fechamento numa decisão provisória com fins operatórios. São os parâmetros de “contrastividade” (externa e interna) do corpus que, no jogo das semelhanças/ diferenças, fazem sentido.

Como fazer, então, entre a concepção de um corpus completamente fechado sobre si mesmo, mas a respeito do qual sabemos que há necessidade de confrontá-lo a outros corpora para significar, e a concepção de um corpus declarado aberto, mas cujos horizontes devem ser delimitados sob pena de não conseguir torná-lo operatório ?

Procedemos, de nossa parte, segundo o método do caramujo : partir de um primeiro corpus nuclear determinado segundo parâmetros de tempo, espaço, gêneros, dispositivos, locutores, temas etc., em função dos objetivos de análise que foram propostos ; para depois estender progressivamente esse corpus confrontando-o com outros, tanto quanto for necessário, em função das questões que surgem à medida que prosseguem as análises.

Daremos, como exemplo, o estudo dos discursos de uma campanha eleitoral. Inicialmente, construímos o corpus das declarações do candidato X do qual fazemos a análise segundo diferentes categorias. Depois, podemos confrontar esses resultados com os corpora das declarações dos candidatos concorrentes para aí distinguir as diferenças e os jogos de remissões uns aos outros, uma espécie de jogo de espelhos, como diz D. Mayaffre ao designá-los pela expressão “corpora reflexivos”.

Podemos também comparar esses resultados aos corpora dos candidatos do mesmo partido das eleições anteriores para distinguir em que esse candidato segue a mesma doutrina de seus predecessores ou em que difere dela, e podemos mesmo ampliar a comparação, incorporando o corpus das declarações e escritos que concernem à ideologia a que pertence o candidato, a fim de avaliar os graus de filiação política ; e pode-se fazer esse mesmo trabalho de confronto para cada um dos outros candidatos.

Podemos também comparar as declarações de um mesmo candidato segundo as situações de comunicação nas quais ele se encontra (comício, entrevistas jornalísticas, declarações na televisão, debates etc.), a fim de verificar se suas estratégias de persuasão são as mesmas ou diferentes em cada um desses casos.

Todo corpus remete a outros corpora num encadeamento de contextos situacionais, paratextuais e interdiscursivos, construindo-se progressivamente sob a forma do que G. Genette chama de uma arquiteitualidade, noção retomada pelos que adotam um tratamento informatizado dos textos. É um jogo de reflexividade entre corpora, que só pode ser descoberto, progressivamente, no momento das análises das questões oriundas do confronto entre esses corpora.

A questão do tratamento quantitativo/qualitativo do corpus

Como é possível notar, compartilhamos do ponto de vista daqueles que dizem que o corpus não é a ferramenta da pesquisa, mas o objeto da pesquisa, pelo menos o objeto que foi construído num corpus em função de pressupostos teóricos e metodológicos. Colocamos, então, a questão do modo de tratamento do corpus, em correlação com a questão da pertinência de uma análise quantitativa comparada à análise qualitativa.

A partir de nossa prática de sujeito analisante e do método de análise que experimentamos nos trabalhos do Centre d'Analyse du Discours sobre corpora midiáticos, concluímos que é preciso discernir o que é da ordem da coleta de dados, da seleção das recorrências e da detecção das condições de comunicação.

A coleta dos dados obedece a uma lógica documental. O objetivo é recolher, descrever e classificar os objetos do mundo e os fenômenos sociais, segundo critérios de forma, sem hipóteses

particulares orientando a pesquisa numa ou noutra direção, e isso com um ideal de exaustividade – o que produz classificações de arquivos diversos, catálogos e bases de dados.

A abordagem é aqui “empírico-indutiva”, apoiando-se numa metodologia tecnicista : os marcadores de descrição e as categorias de classificação são pensados simultaneamente com as características da ferramenta tecnológica (o desempenho da ferramenta pode ter influência sobre os marcadores e as categorias), assim como os modos de representação dessas classificações (árvores, nuvens, estatísticas, tabelas de correlações diversas) e de extrações dos dados para fornecer ao utilizador meios de consultálos. É um método “descritivista”. Um tal catálogo é indispensável a toda pesquisa de corpus, mas não devemos fazer com que esse estágio de coleta de material de análise passe a ser considerado como a própria análise, pois este constitui uma posição anterior à análise.

Foi utilizando a base de dados do acervo do Instituto Nacional do Audiovisual (que tem funções de mediateca, entre outras) sobre a televisão, o rádio e a Web (em elaboração) que a equipe do CAD procedeu a suas análises. O levantamento das recorrências implica que a pesquisa seja orientada. Aqui, contrariamente ao caso precedente, os dados são confrontados a variáveis externas que estão em relação de pertinência com um objetivo de análise, e as recorrências são levantadas e classificadas numa certa ordem em função dessas variáveis. Por exemplo, uma ordem de recorrência de palavras e expressões em relação com um tipo de locutor numa situação dada (um candidato numa campanha eleitoral), ou uma ordem de recorrência temática relacionada a um certo suporte, como por exemplo o levantamento dos temas tratados por um jornal a respeito de um certo acontecimento e comparado aos que foram publicados em outros jornais. Podemos, evidentemente, multiplicar as variáveis, e obteremos um panorama dos resultados estatísticos de ocorrências relacionadas a diversas variáveis externas.

Aqui, podemos dizer que, contrariamente ao caso precedente, trata-se de uma análise quantitativa que fornece preciosas informações para outras análises posteriores, ainda mais porque o desenvolvimento de ferramentas técnicas cada vez mais eficientes (softwares e outros recursos digitais) permitem, atualmente, proceder rapidamente a tais levantamentos.

É nesse gênero de análise quantitativa que se inscrevem muitos pesquisadores que se filiam, cada um à sua maneira, a uma linguística de corpus — de corpora extensos, para dizer melhor, pois não há nenhuma linguística que não seja de corpus. Esses trabalhos trazem resultados preciosos, muitas vezes indispensáveis para fundamentar mais objetivamente certas análises semânticas. Porém, para quem se interessa não somente pelo sentido das palavras mas por seu efeito social em situação de comunicação (as palavras como fator de influência social), é preciso fazer as seguintes observações :

1. a importância e o impacto de uma palavra (isto é, o efeito de sentido) não estão necessariamente ligados à sua recorrência : a expressão “fratura social” empregada por Jacques Chirac em sua campanha presidencial de 1995, as palavras “karcher”⁶ e “racaille”⁷ empregadas por Nicolas Sarkozy, só foram empregadas uma vez por seus autores, e o efeito que produziram se deve, evidentemente, ao sentido que tomaram em seu contexto situacional e interdiscursivo, e também ao fato de que ecoaram nas mídias, logo, em outros corpora. O impacto das palavras não obedece necessariamente a um critério quantitativo. O importante é que se inscrevam numa certa rede discursiva e que façam eco a um sistema de expectativa mais ou menos consciente.
2. o sentido das palavras, como sabemos, é plural. Não somente por sua polissemia, mas também porque adquirem, segundo os contextos linguístico, paratextual e interdiscursivo, o que poderíamos chamar de uma “polidiscursividade”. E isso é muito difícil de medir quantitativamente. Daremos aqui um exemplo : há pouco tempo, por ocasião da última campanha eleitoral, levantamentos lexicométricos puseram em evidência a propensão dos dois principais candidatos, Nicolas Sarkozy e Ségolène Royal, a empregar o pronome de primeira pessoa, “je” (eu, em português). Esse achado forneceu subsídios aos comentaristas para concluírem que essa campanha se caracterizou por uma forte personalização.

Devemos notar, antes de mais nada, que, para poder avaliar essa personalização, seria necessário comparar esses dados com os das campanhas precedentes. Ora, se nos reportarmos à campanha de 1988, podemos observar igualmente um emprego importante do pronome “je”, por parte de alguns candidatos. Isso é bastante natural na medida em que um jogo de alternância “je/nous” (eu/nós) está previsto pelo contrato comunicativo do discurso de campanha eleitoral : o candidato se apresenta ao mesmo tempo como o portador da soberania popular (“nous”), e como um líder diferente dos outros (“je”).

Principalmente, se estudarmos o emprego desses “je”, não apenas em seu contexto linguístico, mas em relação com o ambiente próprio a cada declaração política (paratextual) e com o conjunto dos discursos aos quais se faz referência (interdiscurso), percebemos duas coisas : por um lado, que os “je” de Nicolas Sarkozy e Ségolène Royal obedecem a uma mesma preocupação de se destacar do aparelho de seu partido ; são “je” de diferenciação.

Por outro lado, observamos uma diferença entre os dois : os "je" (mais exatamente os "je veux" – em português, "eu quero") de Nicolas Sarkozy constroem a figura de um chefe poderoso e forte, assumindo todas as responsabilidades e decisões, instituindo-se como homem de ação com um ardor de guerreiro ; em contraste, os "je" de Ségolène Royal constroem a figura de uma mulher combatente animada por uma convicção profunda, testemunha de uma atitude de proteção e de escuta, de sinceridade de sentimentos, de um ardor feito mais de coragem, de tenacidade do que de desejo de poder ; são eles os "je" de apropriação.

Desse ponto de vista, as análises quantitativas, que não devem ser recusadas, não podem trazer os resultados que as análises qualitativas, que levam em consideração as redes de interdiscursividade, poderão fornecer. Será inclusive necessário se abster (visto o entusiasmo atual por essas análises de base numérica) de reificar as análises quantitativas, pois isso levaria à neutralização das qualitativas, o que reduziria consideravelmente o poder explicativo da análise do discurso.

Lembremos que a lógica da pesquisa se baseia numa abordagem "empírico-dedutiva" de idas e vindas entre, de um lado, a determinação dos objetivos de análise e das hipóteses significantes, e de outro, os métodos de observação dos objetos e de coleta de dados segundo uma instrumentação que deve poder dar conta dos objetivos e das hipóteses previamente estabelecidos. É num vaivém entre esses dois polos que se elaboram conceitos, instrumentos de análise e interpretação dos resultados.

Resta-nos remeter o leitor aos estudos empreendidos pelo CAD para ilustrar essa abordagem que articula análise quantitativa e análise qualitativa que resumimos aqui em algumas palavras.

Quando se trata da maneira pela qual a televisão relata acontecimentos de ordem social (uma guerra, uma controvérsia social), devem ser levantados, num primeiro momento, os assuntos difundidos sobre a questão ; para isso, devemos mergulhar na base de dados audiovisuais (na França, do Instituto Nacional do Audiovisual) da qual se extraem os assuntos da questão e com a qual se constitui uma "sub-base" (isso corresponde à fase de levantamento de dados). Depois, procedemos a uma decifração sistemática do material coletado, em função de um certo número de parâmetros externos que remetem à situação de enunciação e à temática.

Assim, foi possível estabelecer, para cada um de nossos estudos : Para o dispositivo enunciativo :

1. o tipo de emissão (telejornal, reportagem, entrevistas, debates etc.) que permite determinar a importância dada por cada estação de TV ao modo de tratamento discursivo (descrição, relato, comentário, argumentação) ;
2. a identidade dos atores envolvidos (político, cientista, jornalista especializado, enviado especial, instituição, cidadão etc.) que permite detectar a legitimidade e a autoridade das intervenções.

Para a temática :

1. o capital temático (em número de sujeitos e em duração de aparições) que cada estação dedica ao acontecimento-tema. Isso fornece um dado objetivo quanto à importância dada ao acontecimento por cada uma, em comparação com a concorrente ; a hierarquização temática (por média horária) relativa ao momento de aparição do acontecimento-tema em cada edição de telejornal (começo, meio, fim), dado objetivo quanto à estratégia de tratamento no interior da programação (o "condutor") de cada estação ;
2. a distribuição dos domínios cênicos, como variável de conteúdo, para determinar a importância respectiva dos universos de referência convocados no tratamento do acontecimento (cenas civil, humanitária, diplomática etc. para a guerra na Bósnia ; cenas científica, política, ética etc. para a controvérsia sobre a clonagem), que apontam para a orientação dramatizante ou argumentativa proposta por cada estação.

Essa fase quantitativa permite, por um lado, constituir índices baseados em resultados estatísticos, índices que são suscetíveis de desempenhar o papel de sintoma, e sobre os quais se desenvolvem análises qualitativas ulteriores ; por outro lado, ela permite constituir um corpus-amostra, isto é, um conjunto de fragmentos de texto que pode ser considerado como representativo em relação a categorias que servirão para analisá-lo de maneira qualitativa : a fala dos atores, as características do dispositivo, o tratamento da temática.

Pode-se escapar assim à crítica que é feita à amostragem, a saber, que é apenas uma parte que pretende representar um todo. Nossa amostra não é a parte de um todo (como seria uma amostra de palavras representativas de um todo textual) ; ela é constituída por categorias em que cada uma é um ponto focal sobre o qual incide a análise qualitativa (os locutores, os gêneros, os temas).

Desse modo, o estudo quantitativo faz sentido em si, mas um sentido provisório que deve ser confirmado, corrigido ou mesmo contradito e, em todo caso, estendido e aprofundado pela análise qualitativa.

Traduzido por :

Angela M. S. Corrêa
Do Departamento de Letras Neolatinas da
Faculdade de Letras da UFRJ e membro do CIAD-Rio

Referências

- ANSCOMBRE, J.-C. et DUCROT, O. L'argumentation dans la langue. Bruxelles : Mardaga, 1983.
- AUTHIER-REVUZ, J. Hétérogénéité montrée et hétérogénéité constitutive : éléments pour une approche de l'autre dans le discours. *Revue DRLAV*, 26, 1982, pp 91-151.
- BARTHES, R. Roland Barthes par Roland Barthes. coll. "Écrivains de toujours". Paris : Seuil, 1975.
- BARTHES, R. S/Z. Paris : Seuil, 1970.
- BOURDIEU, P. Choses dites. Paris : Minuit, 1987.
- BOYER, H. "Solidarité". *Fortune politique, dérive lexicale. Revue Mots*, no 7, Paris, Fondation des sciences politiques, octobre, 1983.
- _____. Sociolinguistique : faire "corpus" de toute(s) voix ?. *Mots*, n°69, Lyon, ENS édition, 97-101, juillet, 2002.
- _____ et LOCHARD, G. Scènes de télévision en banlieue. Paris : Ina-L'Harmattan, 1998.
- BRANCA-ROSSOF, S. Les lettres de réclamation adressées au service de la redevance. *Revue Langage et société* n°81. Paris : Larousse, 1997.
- CHARAUDEAU, P. (dir.) La télévision et la guerre. Déformation ou construction de la réalité ? Le conflit en Bosnie (1990-1994). Bruxelles : Ina-De Boeck, 2001.
- _____. Visées discursives, genres situationnels et construction textuelle. In : CHARAUDEAU, P. *Analyse des discours. Types et genres*. Toulouse : Éd. Universitaires du Sud, 2001, pp.45-73.
- _____. Le contrat de communication dans une perspective langagière : contraintes psychosociales et contraintes discursives. In BROMBERG, M. et TROGNON, A (dir.). *Psychologie sociale et communication*. Paris : Dunod, 2004.
- _____. Le discours politique. Les masques du pouvoir. Paris : Vuibert, 2005.
- _____. (dir.) La médiatisation de la science. Clonage, OGM, manipulations génétiques. Bruxelles : Ina-De Boeck, 2008.
- _____. Entre populisme et peopolisme. Comment Sarkozy a gagné. Paris : Vuibert, 2008.
- _____, LOCHARD, G., SOULAGES, J.-C. La construction thématique du conflit en ex-Yougoslavie. *Revue Mots* n°47. Paris : Fondation des sciences politiques, 1996, pp. 89-107.
- Diadorim, Rio de Janeiro, Volume 10, p. 01 - 23, Dezembro 2011.
- _____ et MAINGUENEAU, D. Dictionnaire d'analyse du discours. Paris : Seuil, 2002 (Em português : Dicionário de análise do discurso. São Paulo : Contexto, 2004. Tradução coordenada por Fabiana Komesu).
- CHAROLLES, M. Cohérence, cohésion et pertinence du discours. *Revue Travaux de linguistique*, n°29, 1995, pp. 125-151.
- COURTINE, J.-J. Quelques problèmes théoriques et méthodologiques en analyse du discours. A propos du discours communiste adressé aux chrétiens. *Revue Langages*, n°62. Paris : Larousse, 1981, pp. 9-128.
- DALBERA, J.-Ph. Le corpus entre données, analyse et théorie. *Revue Corpus* 1, pp 89-101, Faculté des Lettres de Nice 2002.
- DUCROT, O. et alii. Les mots du discours. Paris : Minuit, 1980.
- _____. Dire et ne pas dire. Paris : Hermann, 1991.
- GODIN, C.î. Dictionnaire de philosophie. Paris : Fayard, 2004.
- GROUPE SAINT-CLOUD. Les présidentielles. Regards sur les discours télévisés. Paris : Ina-Nathan, 1995.
- HABERT, B. et alii. Les linguistiques de corpus, Paris : A. Colin, 1997.
- HARRIS, Z.S. Analyse du discours, trad. fr. *Revue Langages*, 13, pp. 8-45 (1ère éd. 1952 : *Discourse analysis, Language*, vol. 28, 1-30), Paris : Larousse, 1969.
- KRIEG-PLANKE, A. La notion de "formule" en analyse du discours. Cadre théorique et méthodologique. Besançon : Presses Universitaires de Franche-Comté, 2009.
- MAINGUENEAU, D. Initiation aux méthodes de l'analyse du discours. Paris : Hachette-Université, 1976.
- _____. Les termes clés de l'analyse du discours., coll. "Mémo", Paris : Seuil, 1996.
- _____ et COSSUTTA, F. L'analyse des discours constituants. *Revue Langages*, n°117, pp. 112-125. Paris, Larousse 1995.
- MAYAFFRE, D. Les corpus réflexifs : entre architextualité et hypertextualité. *Revue Corpus* 1, pp. 51- 69. Faculté des Lettres de Nice, 2002.
- _____. Les corpus politiques : objet, méthode et contenu". *Revue Corpus*, n° 4, pp. 5-19, Faculté des Lettres de Nice.
- MOESCHLER, J. et REBOUL, A. Dictionnaire encyclopédique de Pragmatique. Paris : Seuil, 1994.
- PÊCHEUX, M. Les Vérités de La Palice. Linguistique, sémantique, philosophie. Paris : Maspero, 1975.

Diadorim, Rio de Janeiro, Volume 10, p. 01 - 23, Dezembro 2011. 23

_____. Mise au point et perspectives à propos de l'Analyse Automatique du Discours. Revue Languages, n°37, pp. 7-80. Larousse, Paris, 1975.

QUÉRÉ, L. L'espace public comme forme et comme événement. In : ISAAC, J. (dir.). Recherches. Paris : s.n.b., 1995.

SOULAGES, J.-C. Identités discursives et imaginaires figuratifs. In : HOUDEBINE, A.-M. (dir.), L'imaginaire linguistique. Paris : L'Harmattan, 2003, pp. 103-109.

TOURNIER, M. Verbete "Lexicométrie", IN : Dictionnaire d'analyse du discours, Paris : Seuil, 2002.

VAN DIJK, T.A. et KINTSCH W. Strategies of discourse comprehension. New York : Academic Press, 1983. WILLIAMS, G. (éd.) La linguistique de corpus. Presses Universitaires de Rennes, 2005.

[1] er a Revue Corpus editada pela Universidade de Nice-Sophia Antipolis, especialmente os números 1 (2000) e 4 (2005)

[2] N. T. : documentos constituídos pelos conjuntos de petições e reclamações que cada uma das classes dos "Estados Gerais" da monarquia francesa anotava para serem levados ao rei – os mais famosos foram os redigidos em 1789, por ocasião da reunião dos Estados Gerais que resultou na revolução francesa.
