

Programa Alceste, primeira lição: a perspectiva pragmatista e o método estatístico

Laura Câmara LIMA

Resumo

O objetivo deste artigo é trazer informações a respeito da perspectiva pragmatista e do método estatístico que fundamentam o programa Alceste. Nesse modelo, sujeito, objeto e representação constroem-se uns aos outros, gerando formas lexicais no decorrer do processo. O método permitiu a elaboração de um programa capaz de examinar essas formas lexicais, rastreando os vestígios da atividade dos sujeitos enunciadorees que aparecem, sob a forma de lugares insistentes ou mundos lexicais, em torno de fundos tópicos. Mapeando esses mundos lexicais, o Alceste fornece aos cientistas sociais meios suficientes para que compreendam posições relativas e tensões estabelecidas entre os sujeitos enunciadorees.

Palavras-chave: Método Alceste. Análise pragmática. Mundos lexicais. Sujeito enunciadoree.

Abstract

The objective of this paper is to reveal relevant information about the pragmatist perspective as well as about the statistic method, which underpin the Alceste Program. In accordance with such perspective, subject, object and representation intrinsically interrelate, generating, in turn, lexical forms during this process. This method allowed the lexical forms just mentioned tracking the traces of the activity developed by the subject of enunciation, which emerges as insistent places or lexical worlds, around topic grounds. Through the mapping of these lexical worlds Alceste provides to the social psychologist the necessary means to understand the relative position and the existent tensions between subjects of enunciation.

Keywords: Alceste method. Pragmatic analysis. Topic grounds. Lexical worlds. Subject of enunciation.

Introdução

O programa “Alceste” de análise textual informatizada e automatizada é bastante conhecido de um certo grupo de pesquisadores em Ciências Sociais. São eles franceses, brasileiros e estudiosos de outras procedências, muitos dos quais psicólogos sociais que se servem da Teoria das Representações Sociais e trabalham nos domínios da saúde, da educação, do esporte, da cultura, do lazer, do imaginário, do trabalho, da política.

Apesar de ser mais comumente usado como instrumento analítico no contexto de pesquisas científicas, o programa também pode servir a outros fins, tais como pesquisas de *marketing* e estratégias de comunicação. Ademais, todos os tipos de documentos, provindos das mais variadas origens e tendo os mais diferentes estilos, podem ser analisados sem restrição. Geralmente os textos estudados procedem de livros, dicionários, revistas especializadas, meios de comunicação de massa, entrevistas, testemunhos, listas ou grupamentos de palavras obtidas por associação livre.

Apesar da forte difusão e da boa reputação das quais o programa dispõe, nem todos os usuários conhecem devidamente a teoria semiológica e o método estatístico que o fundamentam, sendo incapazes de compreender mais amplamente as análises feitas por seu intermédio e de interpretar de modo mais consistente os resultados obtidos com o seu emprego. Essa lacuna ocorre em função de os escritos de Max Reinert, o inventor do método, serem pouco divulgados, o que justifica o acesso limitado a tais fontes e o fato de esses mesmos usuários ignorarem que, além de estatístico, ele é também semiólogo, conhecedor de Filosofia e de Psicanálise.

A finalidade deste artigo é estreitar essa carência, explicando os conceitos da pragmática elaborada por Reinert e esclarecendo os fundamentos do seu método, posto em prática através do algoritmo do programa de análise textual que ele próprio batizou de “Alceste” (*Análise dos lexemas co-ocorrentes num conjunto de segmentos de texto*).

Conceitos de base da perspectiva pragmatista

A semiologia¹ desenvolvida por Max Reinert inspira-se principalmente em dois paradigmas teóricos: a Psicanálise de Jacques Lacan (1901-1981) e a Semiótica de Charles Sanders Peirce (1839-1914). A exemplo desses dois, Reinert (2001) também funda a sua perspectiva *pragmatista* sobre um paradigma triádico. Para ele, a noção de *representamem* definida por Peirce (1978, p. 117): “[1.541] [...] o sujeito de uma

1 Neste texto, os termos semiologia e semiótica serão usados como sinônimos, pois uma discussão sobre as definições desses conceitos fugiria aos objetivos do artigo. Ao leitor interessado em aprofundar-se na questão recomenda-se, por exemplo: NÖTH, W. Semiótica e semiologia: os conceitos e as tradições. Com Ciência, Revista Eletrônica de jornalismo científico, localizada na seguinte página web: <<http://www.comciencia.br/comciencia/?section=8&cdicao=11&cid=82>>. Acesso em: abril2008.

relação triádica, com um segundo chamado seu objeto, para um terceiro chamado seu interpretante [...], está ligada à noção de *significante* estabelecida por Lacan, cuja materialidade se manifesta na elaboração que o *sujeito* faz para transformar um *signo* em outro *signo*.

Em seus artigos, Reinert não explicita como os conceitos lacanianos influenciaram suas teorizações, mas menciona duas razões que o levaram a retomar as idéias de Peirce: a) as descrições conceituais da semiótica peirciana têm uma qualidade “profundamente dinâmica”; e b) Peirce coloca em evidência o fato de o *sentido* “original” de um texto (discurso) ser inatingível (e sempre será), porque ele é, ao mesmo tempo, efêmero e mutante: efêmero por durar o ato da sua enunciação e mutante por adquirir outros *sentidos* em função das re-leituras, ao fio das atualizações.

Retomemos, então, num primeiro momento, as definições que Peirce deu aos conceitos chaves *signo*, *representamen* e *semiose*, a fim de melhor entendermos, na seqüência, o paradigma teórico elaborado por Reinert. Para tanto, lançaremos mão de trechos extraídos diretamente da coletânea de textos publicados em francês no livro *Écrits sur le signe* (PEIRCE, 1978) e traduzidos para o português (sob os meus cuidados):

[1.540] Por *signo* eu entendo tudo o que comunica uma noção definida de um objeto, de qualquer maneira que seja, uma vez que essas comunicações de pensamento nos são familiares [...] (p. 116, grifo do autor).

[...]

[2.228] Um *signo* ou *representamen* é alguma coisa que toma o lugar de alguma coisa para alguém, sob alguma relação ou de alguma maneira. Ele se endereça a alguém, isto é, cria no espírito desta pessoa um *signo* equivalente ou talvez um *signo* mais desenvolvido. Este *signo* que ele cria, chamo-o de *interpretante* do primeiro *signo*. Este *signo* toma o lugar de alguma coisa: do seu *objeto*. [...] (p. 215, grifo do autor).

[...]

[5.484] Mas por ‘semiosis’, eu entendo, ao contrário, uma ação ou influência que é ou implica a cooperação de *três* sujeitos, tais que um *signo*, seu objeto e seu interpretante, esta influência tri-relativa, não sendo de maneira nenhuma redutível às ações entre pares. [...] (p. 133, grifo do autor).

Vejamos agora como o próprio semiólogo *pragmatista* desenvolveu as definições dos conceitos de *objeto*, *signo*, *sentido*, *significação* e *semiose*, dentro do quadro teórico que concebeu e operacionalizou com o método Alceste. Lembremos que Reinert (2001a, p. 33, grifo do autor) os formulou de maneira a poder lidar com o caráter efêmero e volúvel do *signo*:

Nesta perspectiva pragmatista, o ‘objeto’ do *signo* não é um referente estático, denotado por um *signo*, separado de uma prática; objeto e *signo* estão envolvidos em um mesmo uso. Assim o objeto não se confunde com a sua aparência num *signo* particular, mas ele é *dinâmico*, do fato mesmo que ele é vivo neste uso: ele se mostra no movimento de um *sentido*. A noção de *signo*

deve então ela mesma incluir o tempo da semiose. Desta maneira, um signo não é simplesmente uma aparência, mas uma certa maneira de passar de uma aparência à outra.

Nessa semiologia, o autor trabalha com as diferentes temporalidades instauradas *pelo e no* discurso e com a multidimensionalidade do “*real*”, partindo do princípio que existem pelo menos três formas de acesso à realidade, definidas por Peirce e relativas às três categorias do conhecimento (os modos de operação do pensamento-signo que se processam na mente): primeiridade, secundidade e terceiridade. Ele considera que cada uma delas está relacionada com a produção de signos de tipos diferentes. Nessa acepção, a construção discursiva subordina-se a uma relação *sujeito-objeto* dinâmica, que mantém a troca simbólica ativa, gerando contínuas transformações nos *sujeitos* e nos *objetos* inter-relacionados.

Retomando o aspecto temporal do discurso, que havia sido apontado por Foucault (1971), Reinert afirma que “O objeto de um discurso está ele também, ao mesmo tempo, na origem do discurso, que ele provoca, e no horizonte de todo enunciado, pois falar dele não esgota o real do objeto” (p. 33). E acrescenta: “É por esta oscilação do sujeito ao seu objeto, do sujeito ao seu vir-a-ser, que o discurso se constrói” (REINERT, 1999, p. 61).

Segundo Reinert (20001a), o *sentido* se estabiliza, se conserva e se impõe por meio do que Peirce denomina *hábito* e reconhece como tendo duas características: uma repetição de formas, que assegura a sedimentação dos signos e lhes dá uma estrutura e uma matéria; e uma dinâmica, que atualiza um fragmento da história do *sujeito* numa situação dada, em relação a um *objeto* determinado.

A experiência do *hábito* constrói a história do sujeito, na medida em que ela o ajuda a se adaptar às várias situações pelas quais vai passando, tecendo as constâncias que asseguram a materialização de sua identidade e a perpetuação de sua transitividade. É a expressão metafórica do *hábito* levando o sujeito a sintetizar dois tempos em um só: reminiscências do passado e projeções do futuro: “É a través da construção de signos que se organiza o jogo mesmo das repetições, inicialmente ligado a impressões múltiplas e passageiras, posteriormente ligado a *habitus* estabilizados, expressões da história e da identidade de um sujeito” (REINERT, 2001b, p. 387, grifo do autor).

Para Reinert (2001a) *significar* é construir planos, roteiros, frases, conceitos, formas, e a *significação* é a parte do *sentido* que se estabiliza, se modela, se calcula; aquilo que um signo, pela sua estrutura, permite representar de outro signo. O teórico vê a *semiose* como uma produção de *sentidos*, em que *sujeito*, *objeto* e *representação* se constroem mutuamente, adquirindo forma na linguagem.

O caráter dinâmico do sentido e a dificuldade de apreendê-lo

Com o propósito de pôr em evidência o caráter dinâmico do *sentido* e a dificuldade de acesso ao momento inaugural de um discurso produzido por um sujeito-enunciador, Reinert (2001a) chama a atenção para o fato de que o ato de representar é paradoxalmente repetitivo e criativo, ao mesmo tempo, implicando numa relação ativa, em que *sujeito* e *objeto* encontram-se interligados um ao outro. Essa característica influenciável e mutante do *sentido* faz com que a tarefa de apreendê-lo, estudá-lo e compreendê-lo torne-se mais laboriosa:

Certo, um signo traduz um outro, para um sujeito, se ele representa um mesmo objeto. Mas a medida deste ‘mesmo’ é toda relativa, e traz a marca da história deste sujeito. Procurando o bom interpretante, o sujeito projeta diante dele, num futuro sempre fugidio, o objeto que o empurra pelas costas impelindo-o a prosseguir a sua busca. O sentido escapa justamente porque o seu objeto é anterior à captação consciente, provocando-a (p. 33).

O *sentido* circula inexoravelmente nas comunicações intergrupais e na medida em que ele se reconstitui (de modo quase idêntico, mas sempre de outra maneira), em cada atualização, em cada tradução de um signo em outro, em cada troca estabelecida entre dois ou mais enunciadores.

Considerando essa dificuldade de se atingir o sentido, Reinert conclui que ele só pode ser abordado por intermédio das “pegadas de passos” que provêm diretamente da atividade coerente de um sujeito-enunciador. Nesses termos, recorrendo a Foucault (1971), diz: “Assim, o tema circula de enunciado em enunciado, dando um sentido ao discurso”. E, sobre isso, faz uso do que dizem Achard (1993) e Wald (1999): “O que resta desta circulação do sentido num discurso é o texto. Mas o *sentido* não está no texto, o *sentido* está no tempo desta circulação, no dinamismo de uma palavra real” (grifo do autor). No entanto, Reinert (p. 33-34, grifo do autor) adverte que restam no texto traços formais da passagem do objeto, não somente através das significações construídas, representadas, mas também através do que se mostra somente como pegadas de passos.

Tais reflexões levaram o pesquisador a conceber um método estatístico de análise de dados textuais, cujo objetivo não se restringe meramente à análise da significação lingüística de um *corpus*, mas se volta ao rastreamento dos vestígios da atividade (da “passagem”) de um sujeito enunciador; os quais são organizados sob a forma de *mundos lexicais*, reunidos nos *fundos tópicos* de um discurso. Nas palavras de Reinert (1993, p. 12),

Dito de outro jeito, eles superpõem, num mesmo ‘lugar’, diferentes momentos da atividade do sujeito, diferentes ‘pontos de vista’. Este lugar atua assim como um ímã em relação a esta atividade. Um sujeito o habita de uma certa

maneira. No caso em que este sujeito é coletivo (caso do estudo apresentado em seguida), estes ‘lugares’ se tornam uns tipos de ‘lugares comuns’ (a um grupo, uma coletividade, uma época etc.).

Uma vez que os “lugares” são habitados por um *sujeito coletivo*, eles ficam impregnados de determinadas características sociais. O “lugar comum” de um grupo ou comunidade releva os hábitos lingüísticos e sociocognitivos partilhados por um ou por outro, e, mediante seu estudo, o pesquisador atinge os nódulos culturais das representações sociais: a cognição partilhada, a experiência conjunta marcada por uma prática “ritualizada” (repetitiva, não necessariamente mística ou religiosa), por sua vez geradora de um léxico próprio, distinguido pela reiteração de certas expressões, combinações lexicais. A relevância em se utilizar o programa Alceste reside na vantagem de ele detectar e identificar, pela via das classes, os lugares do discurso no qual um *sujeito coletivo* se manifesta.

Características do método pragmático

Com respeito à tripartição da teoria da linguagem, a perspectiva de Reinert (1993) é predominantemente pragmática (mais que sintática ou semântica), visto que faz referência à posição do locutor; ao tipo de contexto; aos aspectos performativos dos enunciados, que se sobrepõem aos declarativos; e ao “uso” da língua nas suas dimensões práticas.

Um dos elementos marcantes dessa metodologia é o seu desligamento das noções de sujeito (quem fala) e de ponto de vista ou lugar (de onde se fala), pois um mesmo indivíduo ou sujeito pode falar de vários e diferentes “lugares”, manifestando opiniões diversas, mobilizando distintas maneiras de abordar a realidade. Igualmente, diferentes indivíduos ou grupos de sujeitos, mesmo tendo opiniões contrárias, podem compartilhar um mesmo contexto discursivo, uma mesma maneira de abordar a realidade.

Ainda, pode-se qualificar o método Alceste como pragmático, porque ele utiliza as noções de *sujeito enunciadador* e *enunciado* (preferindo-as às noções de *indivíduo* e de *proposição*), e, para explicitar as conseqüências dessa mudança de perspectiva, assim Reinert trabalha a integração de ambas as noções: o *sujeito-enunciador* é o responsável pela coerência de um “ponto de vista” (munido de sua lógica particular), de um “lugar” relativo (que não existe em si mesmo), “habitado” por um (ou mais) sujeito(s), identificável por sua coerência interna e por sua oposição a outros lugares. Esses *mundos*, por sua vez, formados pela reiteração insistente desses lugares, são mais abrangentes, permanentes e estáveis do que os *pontos de vista* que os compõem.

E o semiólogo diz:

Assim um *ponto de vista* correspondente a uma posição mais consciente e imediata do sujeito-enunciador, enquanto que o mundo do qual este *ponto*

de vista depende, tem uma maior permanência, uma maior estabilidade. Ele é também mais inconsciente (quanto mais ele é pressuposto, mais ele é inconsciente: ver paradigmas de Kuhn) (p. 14, grifo do autor).

Prossegue o autor com sua teoria, descrevendo as três características do *enunciado*, que recupera diretamente da pragmática de Peirce: ser tópico; ser a expressão de um ato intencional particular em situação e de um engajamento ancorado numa temporalidade; e reenviar a uma esquematização daquilo que se fala, num momento do discurso ou do diálogo. Na visão de Reinert, o *enunciado* é uma verdadeira encenação, na qual três elementos atuam: o *mundo lexical*, o *sujeito-enunciador* e a *lógica local*. As relações entre o *sujeito-enunciador* e o *objeto* não somente motivam as mensagens, como também determinam a *lógica local* e a estrutura nela implícita, deixando índices que ficam impressos no texto.

Nossa hipótese é que a ordem do discurso se constrói inicialmente a través de uma dinâmica dialógica, cada *enunciador* só podendo se ancorar nos seu próprio lugar (sua própria história significativa). As relações de ‘força’ entre *enunciadores* se exprimem através da insistência de certos lugares, os lugares estando conotados por tipos de repetições. O objetivo do método é de permitir uma visualização destes lugares os mais insistentes, por intermédio dos *Mundos lexicais*. O objetivo final sendo de visualizar as relações de força e de ajudar a sua superação, pelo fato mesmo que uma representação dos conflitos já é o esboço de uma nova coerência (de uma nova maneira de perceber o estável, o *Mesmo*) (REINERT, 2001b, p. 390, grifo do autor).

Em resumo, a solução que o Alceste propõe é operar uma abordagem *pragmática* do texto centrada na co-ocorrência lexical, na co-presença do léxico numa unidade contextual do texto, diferentemente do que proporia outro tipo de abordagem cuja função recaísse na detecção das frequências de aspectos lexicais tomados isoladamente: “Não se trata de comparar as distribuições estatísticas das ‘palavras’, nos diferentes *corpus* dados, trata-se de estudar a estrutura formal das suas co-ocorrências nos ‘enunciados’ de um *corpus* dado” (REINERT, 1993, p. 9).

O método se distingue, então, por tratar dos usos específicos do léxico, inscrito numa conjuntura discursiva dada, e não de significações estabelecidas a priori pela língua, independentemente do contexto no qual as palavras se encontram, como faria um método puramente semântico. Reinert (1999, p. 68, grifo do autor) explica: “Em resumo, no nosso modelo, nós não procuramos mais representar a forma lógica de uma representação subjacente a um conjunto de proposições, mas a ‘cartografar’ os principais *topoi* ou ‘lugares comuns’ sobre os quais este mundo do discurso se constrói simultaneamente com aquele dos enunciadores”.

Concluindo, o que o Alceste examina no texto não são as divisões superficiais provenientes das significações de palavras isoladas, mas, ao contrário, as ressonâncias de *sentido* que se estabelecem devido às co-ocorrências de alguns termos e que aparecem reunidas em certas regiões do texto (discurso) em certos momentos. Logo, esse

método não demanda uma compreensão ativa, que parta das significações “em-si” do léxico, extraída daquela pequena lista de significações gerais citadas no dicionário e que servem de definição ou de aceção do conceito. Ele nos convida a seguir os caminhos oferecidos, percorrendo as trilhas gravadas no relevo do texto, levando-nos a compreender por quais veredas de *sentido* esses agrupamentos lexicais nos conduzem.

Nas palavras de Reinert (2001a, p. 34),

Nosso intuito é, sobretudo, partir no sentido contrário, em direção desta origem objetiva e dinâmica presente nos traços mais imediatos, menos pensados. E para isto é preciso desmontar o texto; a cada passo, detectar uma marca do que se oferece espontaneamente. Para isto, o discurso deve ser entendido no ritmo dos seus momentos sempre renovados. Não é tanto quem fala, ou o que se diz, que nos interessa, mas de onde se fala a cada instante!

A diferença entre a análise Alceste e a análise de conteúdo

O equívoco mais comum associado ao método Alceste consiste em considerar que ele opera uma análise de conteúdo. Essa visão limitada das propriedades do programa foge ao que Reinert (2001a, p. 33) propõe: “O método que propomos reconcilia de uma certa maneira, o objeto de uma análise de conteúdo com o sujeito de uma análise de discurso, como dois momentos se articulando no percurso de uma leitura”.

As classes identificadas pelo programa são incorretamente concebidas como categorias puramente semânticas, desligadas dos *fundos tópicos* e das forças dinâmicas que as originaram. Enquanto isso, na proposição do criador do programa elas podem (e devem) ser interpretadas sob três pontos de vista:

- a) como *conteúdo*, observando-se a lista de palavras ou de unidades de contexto elementar (u.c.e.) que lhes são específicas;
- b) como *atividade*, observando-se as passagens de uma classe a outra no decorrer do processo discursivo; e
- c) como *representação*, observando-se que as classes formam um sistema e refletem certa estabilização da atividade do sujeito enunciatador (2001a p. 34).

A fim de elucidar as diferenças entre os dois métodos, tratemos de recapitular as limitações da análise de conteúdo e as especificidades da análise do Alceste.

No livro de Bardin (1977), classicamente usado como referência no assunto, a análise de conteúdo é definida segundo a aceção de Berelson (1971), como uma técnica de pesquisa para a descrição objetiva, sistemática e quantitativa do conteúdo manifesto das comunicações, tendo por objetivo de interpretá-las. O método é bem conhecido e consiste na tomada de consciência dos conteúdos presentes no

texto analisado; na construção de categorias em que eles possam ser classificados; na distribuição dos conteúdos em função das categorias previamente nomeadas; na elaboração de comentários e interpretações a partir das características de cada uma das classes.

A limitação desse tipo de procedimento incide sobre o resultado a se alcançar (ou seja: a classificação feita), dependendo do pressuposto do qual se parte (ou seja: as categorias). Seu ponto fraco refere-se ao fato de que o critério usado para o estabelecimento das categorias impõe necessariamente um corte de realidade, variando segundo os indivíduos. A experiência mostra que dificilmente os quesitos utilizados na classificação de objetos e/ou palavras como categorias e subcategorias são os mesmos e que raramente as categorias criadas pelos sujeitos são equilibradas e subdivididas da mesma maneira. Sendo assim, o conteúdo encontrado dependerá do tipo de categorização que o pesquisador utiliza.

Por outro lado, a forma pela qual o programa Alceste obtém as classes é diferente, pois a classificação não é feita em relação ao significado do léxico e sim em relação ao uso que deles se faz. Não é o pesquisador que define o número e o critério de separação das classes, mas é o programa que descobre quantas classes estáveis o texto contém a partir de certa fragmentação ideal, obtida por meio de testes sucessivos (que explicaremos a seguir).

Para Reinert (2001a, p. 34, grifo do autor), “É uma outra maneira, diferente da maneira da análise de conteúdo, de falar da *origem objetiva* do sentido num discurso. Nós não damos, ainda mais, nenhum conteúdo particular, à priori, aos nossos fragmentos de texto, que não tem que circunscrever nenhuma significação. Eles só delimitam momentos na atividade discursiva”.

A vantagem desse método recai sobre o fato de ele ser capaz de captar condensações de *sentido* próprias e específicas do texto, que a racionalidade das análises de conteúdo deixa passar despercebidas. E isso porque a formação de sentido, que é própria de um texto determinado, diz menos respeito aos conteúdos associativos das enunciações que a um processo inconsciente referentes às posições relativas dos enunciadores. Enfim, por essa via de análise, o autor propõe captar o *sentido* na sua dupla riqueza: na “fibra do real”, como percurso, como sucessão de momentos; e na “fibra do imaginário”, como marca de uma ancoragem intuitiva, lugar, matéria de uma crença do sujeito em seu mundo.

Os conceitos fundamentais do método Alceste

Essas reflexões semiológicas e metodológicas levaram Max Reinert a conceber um programa informático por ele batizado de Alceste, cujos conceitos fundamentais são três: *palavras plenas*, *mundos lexicais* e *fundos tópicos*. Seguem-se as definições.

Conceitualmente, quando Reinert (2001a) usa a expressão *palavra plena*, ele se refere ao “primeiro momento da semiose onde o signo se dá como abundância”, so-

bre o que explica: “Uma palavra plena nos introduz, naturalmente, num lugar-fonte, portador de imagens e afetos [...] Uma palavra é plena, para um sujeito, aqui e agora, porque ela interfere com a sua história [...] e ela é raramente plena sozinha; esta abundância contamina todo um momento [...]” (p. 34).

Desse modo, o programa classifica como *plenos* (e utiliza essa classificação nos cálculos estatísticos) os verbos, nomes, adjetivos, advérbios, cores, meses do ano, dias da semana, lugares, países e nomes ligados à família (além das formas conhecidas não codificadas e aquelas não conhecidas). São, pois, vocábulos que se referem a *sentidos* bem-estabelecidos na língua, ancorados na percepção primeira, na impressão direta do mundo dos sentidos, dos hábitos e da memória, que se impõem ao sujeito. Complementando, o teórico estabelece, então, os termos que ele designa de *ferramenta*, cujo *sentido* se subordina ao *sentido* das primeiras. São os artigos, pronomes (demonstrativos, indefinidos, relativos), preposições, conjunções, interjeições, números, verbos modais, verbos auxiliares, marcadores (de modo, tempo, espaço, intensidade, relação, pessoa), formas não conhecidas frequentes e palavras grafadas com inicial em maiúscula. Trata-se, assim, de palavras cujo *sentido* é sempre relativo, dependente das outras palavras (*plenas* e *ferramentas*) com as quais estão associadas.

Igualmente quanto ao aspecto conceitual e terminológico, Reinert (1993, p. 13, grifo do autor) define *mundos lexicais* como “espaços de referência definidos estatisticamente, associados a um grande número de enunciados”, esclarecendo que “Um mundo lexical é, ao mesmo tempo, o vestígio de um lugar referencial e o índice de uma forma de coerência ligada à atividade específica do sujeito enunciador que chamaremos de *lógica local*”.

Quando Reinert usa o termo *fundos tópicos* ou *fundamentos tópicos*, refere-se a posicionamentos que relevam parâmetros tópicos e analógicos. São “lugares” primeiros, arcaicos, ligados às origens do *sentido*. A dinâmica dos *fundos tópicos* é fruto dos movimentos de alternância entre orientações antagonistas que se reiteram no discurso, e ela apresenta um caráter social, pois pressupõe, necessariamente, a existência de diferentes grupos ou categorias de sujeitos enunciadores.

As relações intergrupais que motivam as enunciações determinam também sua estrutura implícita. Essa isotopia entre forma e conteúdo é o sinal de um *campo contextual*, o índice da presença de um *fundo tópico*. Reinert (2001a, p. 34, grifo do autor) resume, assim, a maneira pela qual articula esses conceitos: “Nossa hipótese é que, em todo enunciado, opera um *fundo associativo*, que se revela através da co-ocorrência das suas palavras plenas. Mais que um *fundo associativo*, nós o chamamos de *fundo tópico*, pois a associação se dá a priori, como vestígio arcaico de uma mesma origem tópica”.

As operações feitas pelo algoritmo Alceste

Concebido para detectar os *hábitos* (PIERCE, 1978) que os enunciadores estabeleceram ao longo do exercício das trocas discursivas que eles entretêm, o programa Alceste trabalha essencialmente com a classificação das configurações lexicais.

Num primeiro momento, o algoritmo reduz as palavras a suas raízes, ou seja, substitui uma forma textual por seu *lema* (na sua forma *standard*, de acordo com os dicionários da língua). Essa redução tem por objetivo melhorar a análise estatística e a classificação das unidades de contexto, reunindo sob uma mesma forma reduzida formas que veiculam uma mesma referência. Esse procedimento é auxiliar no desprendimento dos casos particulares, favorecendo a concentração nas características comuns aos vários léxicos, a fim de se exacerbarem as redundâncias.

Em seguida, os lexemas reduzidos e reunidos em torno das “*raízes comuns*” são separados em dois grupos, *palavras plenas* e *palavras ferramentas*, cada qual tratado de modo diferente: as primeiras são computadas e incluídas nos cálculos; as últimas não entram nesse cômputo, mas são projetadas relativamente aos fatores obtidos na análise daquelas, às quais estão subordinadas. Esse segundo nível de aproximação garante uma informação complementar, em que as *palavras ferramentas* servem de indícios das especificidades de cada um dos discursos; elas dão a “*coloração*” pessoal (ou impessoal) de cada um dos agrupamentos de *palavras plenas*.

A partir do momento em que o algoritmo começa o tratamento das *palavras plenas*, o texto sob análise é dividido em fragmentos, que Reinert batizou de “*unidades de contexto*”, cujo *primeiro recorte* se dá em função dos sinais próprios da língua. Provindas da organização inicial do texto; caracterizadas pela organização das pausas, dos períodos, dos parágrafos; delimitadas pela pontuação e pelos intervalos; essas primeiras unidades são nomeadas “*unidades de contexto inicial*” (u.c.i.).

A esse primeiro seccionamento o programa sobrepõe um *segundo*, baseado em outro critério de divisão textual, então repartido segundo unidades de análise mais reduzidas ainda e que sejam liberadas da organização inicial do texto, de maneira a captar formas de composição lexical as mais elementares. Para isso, o programa usa como critério de demarcação um *número X* de *palavras plenas*; em seguida faz uma segunda série de análises, lançando mão de um número $X+2$ de *palavras plenas*; e por último verifica se os resultados são os mesmos nos dois ensaios. Se tal é o caso, a conclusão é que a análise é boa, por ser estável; se não, o pesquisador deve ir testando sistematicamente todos os tamanhos de unidades ($X-2$ e X , $X-4$ e $X-2$, ou $X+2$ e $X+4$ etc.), até obter classes *estáveis*, ou melhor, aquelas cujo número e disposição sejam sempre os mesmos, independentemente de uma pequena variação no tamanho das unidades utilizadas. Resumindo, o recorte final do texto obedece a dois parâmetros: as pontuações encontradas e certo número (a definir) de *palavras plenas* contadas. Com essa dupla fragmentação, Reinert define o que ele chamou de “*unidades de contexto elementar*” (u.c.e.), a partir das quais o restante dos cálcu-

los é realizado. Trata-se da menor unidade estatística definida pelo programa, em que as *unidades de contexto iniciais* são as maiores.

A próxima operação do programa é montar uma tabela lógica de duas entradas. As *unidades de contexto elementar* obtidas (que correspondem teoricamente aos *enunciados*), numeradas segundo a ordem de aparição no texto, são dispostas em linhas, e as raízes das *palavras plenas* retidas (léxicos) são organizadas em colunas. As casas da tabela, resultantes do cruzamento das linhas e das colunas, são marcadas segundo um código binário. Quando assinaladas com o número “um”, isso indica a presença do léxico na u.c.e., e quando marcadas com o “zero”, significa a ausência do léxico nessa unidade. Se as ausências são mais freqüentes que as presenças, a tabela possui mais “zeros” do que “um”, o que permite ao algoritmo detectar zonas de maior densidade de “um” e zonas vazias. A partir desse material em que se exhibe a presença-ausência, o programa opera uma classificação descendente hierárquica (CDH), técnica descritiva de análise de dados elaborada para tratar tabelas lógicas de grande dimensão, mas de pequeno efetivo. A tabela é decomposta em subtabelas, de maneira a maximizar a quantidade de “um” em cada uma delas, com a menor intersecção possível de vocabulário, as quais, porém, são analisadas da mesma maneira e sucessivamente, cujo encadeamento produz uma árvore hierárquica.

Enfim, o algoritmo isola os fatores que exprimem a separação das classes e apresenta o plano fatorial que resume o número e a distribuição dessas classes (ou *mundos lexicais* ou *fundamentos tópicos*). O mais interessante é que o plano fatorial e a topologia dos fatores sintetizam as polémicas que mobilizam as relações entre os grupos ou as categorias implicadas na problemática tratada.

Para dar conta da dimensão social, o programa ainda trabalha com uma terceira fonte de informação, que ele projeta em relação aos fatores obtidos com a análise das *palavras plenas*. Trata-se das variáveis (e de suas modalidades) estabelecidas pelo pesquisador de acordo com os critérios relativos à perspectiva adotada, as quais trazem informações sociológicas e psicossociais a serem associadas aos *mundos lexicais* detectados, permitindo ao investigador compreendê-los e explicá-los em relação às dinâmicas intergrupais que animam a produção discursiva analisada. Essas variáveis podem ser do tipo sociológico – gênero, idade, formação, profissão, função; do tipo psicossocial – geração, identidade pessoal, identidade profissional, equipes de trabalho, categorias de opinião, tipo de conhecimento coletivo, posição social; ou, ainda, de outros tipos, consoante as linhas das pesquisas.

Considerações finais

O algoritmo Alceste permite a classificação de fragmentos de texto segundo um procedimento que desvenda as utilizações mais contrastadas das *palavras plenas* ali contidas. O vocabulário específico de uma classe caracteriza um tipo de contexto, chamado por Reinert (1990, p. 45) de “campo contextual” (espaço semântico particular, caracterizado com a ajuda de uma classe de palavras), que depende da maneira específica pela qual um *corpus* particular é constituído e das suas leis de produção.

Para o autor, essas *palavras* se situam numa mesma vizinhança temporal, porque compartilham uma mesma origem tópica, motivo pelo qual ele não se interessa em comparar a distribuição estatística das diferentes *palavras plenas* independentemente umas das outras, preferindo se concentrar na estrutura formal de suas co-ocorrências nas diferentes *unidades de contexto* de um texto dado. Essa abordagem rende grande importância para o método, pois permite ao usuário se colocar na perspectiva da abordagem estatística de “contextos tipos”, indo além da abordagem semântica de “contextos locais”. Esses “contextos tipos” são as poucas classes de *palavras plenas* em um *corpus* marcado por certo tipo de posição assumida pelos sujeitos. Nas próprias palavras de Reinert (1990):

As ligações que nós queremos estabelecer entre *unidades de contexto* devem permitir apreender, depois de uma análise estatística, aquilo que nós chamamos de ‘*contextos tipos*’, ao menos os mais impregnados entre eles num *corpus* dado (p. 29, grifo do autor).

[...]

Um *contexto tipo* servirá então para apreender um tipo de mundo; o termo ‘mundo’ é mais conveniente, na nossa opinião, que o termo ‘representação’, pela não diferenciação que ele supõe, traduzindo bem o aspecto noturno, inconsciente desta forma de representação [...] (p. 31-32, grifo do autor).

O que caracteriza o método Alceste como pragmático são o tipo de tratamento das palavras (léxicos) e o tipo de recorte das unidades do texto, divisão essa que não obedece a uma regra semântica, mas a uma regra formal arbitrária, que provoca os cortes a partir de dois critérios combinados: as subdivisões iniciais do texto e a contagem das *palavras plenas*.

Reinert (2003, p. 403, grifo do autor) encontrou na metáfora do grão fotográfico uma ilustração de como segmentos de texto cortados arbitrariamente levam à captação dos *fundos tópicos*:

O grão da foto é definido de maneira ainda mais arbitrária que os *cortes* que nos propomos aos enunciados. E isto importa pouco, na medida em que o grão permite ao *sujeito* encontrar uma *estabilidade* nas *formas* da sua experiência (*significante*). A definição do grão é arbitrária e, no entanto, a possibilidade de *representar alguma coisa* depende dele. No nosso modelo, os grãos são os *segmentos de texto* ou *unidades de contexto* e a coloração dos grãos depende dos *mundos lexicais*.

Com esse método, os signos não são interpretados, a priori, a partir de categorias impostas, de significações convencionais ou gerais. Ao contrário, são considerados segundo a utilização que deles se faz, revelada pelos encadeamentos lexicais, pela significação particular que as palavras tomam num texto específico, pelo vocabulário que o(s) sujeito(s) enunciador(es) utiliza(m). Breve, esse funcionamento permite ao programa seguir os vestígios da(s) lógica(s), da(s) subjetividade(s) do(s) desejo(s), do(s) enunciador(es) ou do(s) autor(es) que produziu(ram) o discurso, atingindo, assim, a dimensão pragmática do texto.

Ora, é justamente a maneira pela qual o programa opera a incisão das unidades do texto que lhe garante originalidade: o Alceste capta sobretudo as rupturas. A consequência é que a repartição das classes não é obra da manipulação do criador do programa ou da categorização imposta pelo pesquisador que o utiliza, mas é fruto da topologia natural do texto, que o algoritmo do programa é capaz de detectar. Ao usuário cabe, então, a tarefa de explicar os fatores (muitas vezes implícitos) que determinaram a repartição das classes que ele encontrou, de modo a compreender as forças implícitas responsáveis pela divisão (reagrupamento) dos vocábulos tal como se encontra). É, pois, função do pesquisador de rastrear os vestígios das pegadas de passos deixadas pelo(s) sujeito(s) enunciador(es), a fim de encontrar o “elo perdido” que o conecte aos sentidos que pairam naqueles mundos lexicais.

Referências

- ACHARD, Pierre. **La sociologie du Langage**. Paris: Presses Universitaires de France, 1993. 128 p.
- BARDIN, Laurence. **Análise de Conteúdo**. Lisboa: Edições 70, 1977. 296 p.
- BERELSON. **Content analyses in communication research**. NY, Illinois University Press. 1959. Hasner, 1971.
- FOUCAULT, Michel. **L'ordre du discours**. Paris: Gallimard, 1971. 81 p.
- NÖTH, Winfried. Semiótica e semiologia: os conceitos e as tradições. **ComCiência, Revista Eletrônica de jornalismo científico**, n. 74, 2006. Disponível em: <http://www.comciencia.br/comciencia/?section=8&edicao=11&cid=82>.
- PEIRCE, Charles Sanders. **Écrits sur le signe**. Paris: Seuil, 1978. 269 p.
- POTTIER, Bernard. **Linguistique générale**. Théorie et description. Paris: Hachette, 1974.
- REINERT, Max. Alceste, une méthodologie d'analyse des données textuelles et une application: Aurélia de Gérard de Nerval. **Bulletin de Méthodologie Sociologique**, Paris, n. 26, p. 24-54, 1990.
- _____. Les «mondes lexicaux» et leur «logique» à travers l'analyse statistique d'un corpus de récits de cauchemars. **Langage & Société**, Paris, n. 66, p. 5-39, 1993.
- _____. Alceste, une méthode statistique et sémiotique d'analyse de discours. Application aux «Rêveries du promeneur solitaire». **La Revue Française de Psychiatrie et de Psychologie Médicale**, Paris, tome V, n. 49, p. 32-36, Outubro, 2001a.
- _____. Processus catégorique et co-construction des sujets et des mondes à travers l'analyse statistique de différents corpus. In: ARRIVE, Michel; NORMAND, Claudine (Dir.). **Linguistique et Psychanalyse**. Paris: In Presse, 2001b. p. 385-398.
- _____. Le rôle de la répétition dans la représentation du sens et son approche statistique par la méthode «ALCESTE». **Semiotica**, Berlin, NY, issue 147, November, p. 389-420, 2003.
- WALD, Paul. **Classes d'énoncés, dimensions modales et catégories sociales dans ALCESTE**. Paris: L'Harmattan (Utinam), 1999.

Recebimento em:	01/02/2008
Aceite em:	04/03/2008