



Diálogos entre a tecnologia, a estatística e a semiótica para uma abordagem de textos

Márcia Angélica dos Santos*

Resumo: O objetivo deste trabalho é verificar a possibilidade de empregar os índices qualitativos e também quantitativos, gerados pela Tabela de Discriminação de Valores Lexicais, como banco de dados para o desenvolvimento de outros métodos. Essa tabela é parte do instrumental metodológico criado por André Camlong (1996), da Universidade de Toulouse II, instrumental este voltado exclusivamente para a descrição e a análise de léxicos, de textos e de discursos. Suas informações são produzidas por operações matemático-estatísticas, executadas com a ajuda da tecnologia computacional. Ao transformar os textos em listas de vocábulos devidamente mensurados, o instrumental disponibiliza essas informações para outros procedimentos, independentes do processo relacionado ao método. Para averiguar essa abrangência de aplicação dos dados, considera-se a Tabela de Discriminação de Valores Lexicais como subsídio do percurso gerativo de sentido, proposto por Greimas e Courtés (1983), com seus níveis: fundamental, narrativo e discursivo. Partindo de articulações sincréticas — de um lado a orientação de dados estatísticos produzidos pela tecnologia; de outro, o percurso gerativo de sentido em sua forma mais tradicional —, a investigação conclui que os itens lexicais, organizados em faixas segundo a frequência e o valor adquirido no ambiente de que fazem parte, são suficientes para subsidiar a proposta semiótica de análise de textos, estabelecendo, nessa combinação, a segurança da descrição científica e atestando sua validade no desdobramento de qualquer metodologia.

Palavras-chave: computação, estatística, léxico

Introdução

A introdução dos avanços tecnológicos nas mais variadas formas de manifestação cultural obteve, como todo processo inovador, a aceitação, marcada pelo entusiasmo excessivo, e a recusa, carregada do repúdio próprio ao conservadorismo, ambos indicativos da necessidade de uma análise mais substancial em relação às possibilidades oferecidas. Passados os primeiros momentos típicos de uma fase de ajustes, a tecnologia vem encontrando seu lugar, ao lado de outras contribuições que igualmente servem ao homem para entender melhor o mundo e o segmento no qual se acha integrado. A tecnologia computacional, por exemplo, inseriu-se em todos os campos do conhecimento e, nos estudos linguísticos e semióticos, não poderia ser diferente. Tendo aí se instalado já há algum tempo, os procedimentos relacionados a computadores progredem no sentido de serem vistos dentro de parâmetros mais objetivos, ao conferir novos formatos às pesquisas e ampliar as propostas interdisciplinares já existentes, tanto na linguística quanto na semiótica. Na análise

de textos e de discursos, as possibilidades de emprego dos computadores distinguem-se quanto às vantagens que oferecem, quando analisadas do ponto de vista utilitário: baseadas em *softwares*, as iniciativas ajudam a listar, contar e ordenar rapidamente diversos aspectos de uma desconstrução e de uma reconstrução de textos, assegurando resultados de procedimentos marcados pela minúcia vocabular, pelo detalhamento sintagmático e pelas ligações das mínimas partes com o todo. No que se refere à análise de texto e de discurso, algumas rotinas evidenciam de imediato certos comportamentos lexicais e discursivos, fato ao qual se juntam as muitas surpresas com os resultados: expressões por vezes minimizadas em abordagens mais ortodoxas, atreladas ao plano da leitura natural repetitiva, revelam-se importantes para a composição considerada. Esses fatores tornam-se mais evidentes com a utilização do método fundamentado no *Stablex*, instrumental criado por André Camlong (1996), pesquisador do Laboratório de Inteligência Artificial da Universidade de Toulouse II.

Nestes anos em que a tecnologia vem interagindo

* Universidade Federal do Tocantins (UFT). Endereço para correspondência: (arkabbys@yahoo.com.br).

com os estudos linguísticos e semióticos, o *Stablex*, cujo nome traz associados *sta* de estatística, *table* de tabela e *lex* de léxico, destaca-se por ser o suporte de um método específico para textos; juntos, instrumental e método deixam visíveis os procedimentos pautados por uma intercomunicação entre diferentes pólos, pois baseiam-se em tópicos matemático-estatísticos para construir, a partir da desconstrução, um banco de informações com propriedades ao mesmo tempo quantitativas e qualitativas. Sua aplicação vale-se de macros e de *softwares* como o Excel, operados em microcomputadores PCs, com o intuito de estabelecer uma sequência mecânica de decomposição imediata dos textos em itens lexicais, acompanhados de valores quantitativos a princípio e quantiquantitativos a seguir: o método, entre outras diretrizes, coloca em evidência, pela análise matemático-estatística, um valor de uso determinante da importância de cada vocábulo na articulação discursiva. Desse modo, os elementos mínimos organizam-se de acordo com certas preferências na articulação do texto, preferências essas que apontam a tendência do sentido. O método é descritivo, objetivo e indutivo: o texto, e tão somente o texto, é o elemento considerado para a rotina que, por meio da informática, processa, armazena e recupera dados, estabelecendo a relação entre valores numéricos e ocorrências lexicais. As tabelas e os gráficos derivados revelam as tendências, agrupando-as em categorias. Os valores quantitativos iniciais, transformados em valores quantiquantitativos pelos cálculos algébricos que fundamentam algumas tabelas, distribuem os vocábulos por zonas de preferência, de base e de rejeição; essa classificação leva a identificar o tema e, por conseguinte, a orientação do discurso.

A abordagem com base nos dados fornecidos pelos módulos do *Stablex* coloca em evidência o comportamento dos textos e pode subsidiar os mais variados objetivos, permitindo trabalhar com qualquer tipo e gênero de texto e de registro linguístico, sejam os da língua materna, sejam os de outros idiomas. Seu banco de dados é, como todo banco de dados, um complexo de informações disponíveis para os mais diferentes fins, o que inclui a possibilidade de estabelecer relações com outros métodos de abordagem de textos: há procedimentos que podem ser desenvolvidos a partir das informações organizadas nas tabelas e nos gráficos, e um deles pode se valer do percurso gerativo de sentido proposto por Greimas e Courtés (1983). No método greimasiano, o texto é a unidade de manifestação composta por três níveis – fundamental, narrativo e discursivo –, todos eles marcados por uma sintaxe e uma semântica próprias. Comparando-se as metodologias, observa-se que o desenvolvimento da proposta greimasiana pode ter como subsídio o léxico dos textos organizado por valores, especificamente os valores fornecidos pela Tabela de Discriminação de

Valores Lexicais (TDVL), também denominada Léxico Preferencial ou Tabela de Valores Lexicais (Zapparoli, 2002). Essa tabela, constituída por valores quantiquantitativos, organiza o léxico de cada texto de modo decrescente, segundo o peso (valor) individual do item na articulação desse mesmo texto, que o método denomina *variável*. A proposta de Camlong e o percurso estabelecido por Greimas possuem procedimentos semelhantes à medida que partem de elementos mínimos geradores. No método Camlong, os itens lexicais, distribuídos segundo o valor no texto considerado e no *corpus*, direcionam o sentido do discurso. Para juntar esses dois procedimentos, é suficiente tomar os itens lexicais destacados pela Tabela de Discriminação de Valores Lexicais como elementos formadores dos níveis instituídos pela abordagem semiótica de Greimas, lembrando que ambos os métodos desenvolvem um comportamento comum: todas as observações são geradas a partir dos textos.

1. Proposição

O método Camlong desenvolve seus procedimentos baseados em diferentes tabelas, tais como Alfa, Delta, de Distribuição de Frequências, de Desvios Reduzidos, de Discriminação de Valores Lexicais, Binomial, de Correlação. No roteiro que aqui se propõe, a totalidade do método não é considerada. Em seu desenvolvimento original, este trabalho observou as determinações de reunir o mínimo de três textos de um mesmo autor, selecionando seis contos da obra de Guimarães Rosa para o desenvolvimento da proposta. Aqui, pretende-se uma demonstração sucinta dessa proposição inicial e, para isso, será focalizado somente o conto “Hiato”, no entanto, todas as considerações fazem parte dessa abordagem mais extensa, que abarca os seis textos roseanos. Também é preciso assinalar que, para utilizar os pesos atribuídos pela Tabela de Discriminação de Valores aos itens lexicais, há necessidade de verificar a homogeneidade do *corpus* escolhido, o que significa testar o equilíbrio lexical dos textos, traduzido pelos índices da Tabela de Desvios Reduzidos (TDR) e pela Tabela de Fisher. Esse procedimento não será demonstrado aqui, devido à sua extensão. O que se tem como foco é tão somente apresentar, por meio de um texto extraído de um *corpus* mais abrangente, a validade da intersecção dos métodos. Ou seja, demonstrar-se-á que as informações quantiquantitativas da Tabela de Discriminação de Valores Lexicais constituem um banco de dados privilegiado para elaborar uma análise pautada no percurso gerativo de sentido.

2. Fundamentação teórica

A Tabela de Discriminação de Valores Lexicais do método Camlong (1996), que também será indicada com a sigla TDVL, é a base dos procedimentos desenvolvidos, pois traz todas as informações necessárias: indivi-

dualiza e identifica os itens lexicais de cada variável, informa a frequência e o peso de cada vocábulo nas relações que mantêm na variável de que faz parte e no *corpus* escolhido, compondo uma hierarquia lexical e uma organização segundo esses valores. Um item não é considerado pelo número de vezes que ocorre e, sim, pelo valor que adquire em relação a outros itens lexicais, no texto e no *corpus* de que faz parte. Orientando quanto ao uso privilegiado ou não de cada item na articulação dos textos e dos discursos, a TDVL é a escolha natural para uma relação com outras propostas voltadas para o sentido. Na abordagem semiótica com base em Greimas, o roteiro é marcado pelos itens lexicais mínimos, em relação de contraste, contradição ou complementaridade; pelos estados e transformações de estados, expressos em programas narrativos com as consequentes destinações e manipulações; pelas debreagens da enunciação e por figuras e temas que constroem o simulacro do mundo, que aqui nesta abordagem recebem um acento da simbologia antropológica de Gilbert Durand (1989).

Para o exame do percurso gerativo de sentido, utilizam-se, como base, sobretudo as obras de Fiorin (1989/2002), Barros (1988/1990) e Greimas e Courtés (1983).

De modo geral, a posição adotada situa-se na estatística e na semiótica, guiando-se pelo princípio que tem no texto o seu limite, ou seja, as informações geradas pelo texto é que subsidiam todos os comentários.

3. Corpus (ou amostragem)

3.1. Seleção

Segundo o método Camlong, recomenda-se o mínimo de três variáveis para a obtenção de dados significa-

tivos. No trabalho original foram selecionados seis contos da obra de Guimarães Rosa; aqui, para a demonstração das possibilidades, focaliza-se um desses contos — “Hiato” — pertencente a *Tutameia (Terceiras estórias, 1985)*. Atendendo também a uma das orientações do método Camlong, o conto passa a chamar-se variável, cujas características básicas abaixo mencionadas podem ser conferidas no desenvolvimento da descrição e da análise.

Delimitou-se a aplicação do procedimento aos textos em que o enunciador, realizando debreagens e embreagens, privilegia um narrador também actante, que, por sua vez, é revestido por um ator, interlocutor ou não. Na variável “Hiato”, ao narrador actante se superpõe, no nível discursivo, a figura de um vaqueiro. É preciso esclarecer, no entanto, que essa caracterização é inferida, já que não é declarada pelo narrador. Em tom de camaradagem, esse ator vaqueiro reflete e estimula a reflexão sobre as crenças e a impotência do saber acumulado pelo homem. Seu trânsito acontece no ambiente rural, uma vez que há citação dos espaços “rincão” e “rancharia” (Rosa, 1985, p. 75, parágrafo 24), e o nome da região “Cambaúba” (Rosa, 1985, p. 72, parágrafo 1)¹ situa esses locais em Minas Gerais. O relato traz a oscilação entre o momento presente da emissão e o pretérito dos acontecimentos relatados.

Tabela de discriminação de valores lexicais (TDVL)

A variável possui 574 vocábulos, que a TDVL organiza segundo o peso que obtêm quando relacionados ao texto e ao *corpus* considerado. Os itens lexicais são elencados segundo esse peso, do maior para o menor, o que traduz a importância de cada um na articulação discursiva. A tabela apresenta, então, uma primeira faixa com o *léxico preferencial*, ou seja, os vocábulos com valor substancial, acima de 1,96.

Ordem	Léxico	Total	Hiato	Valor
1	nhácio	10	10	9,246
2	põe-põe	9	9	8,772
3	vaqueiro	6	6	7,162
4	touro	5	5	6,538
5	nos	18	8	4,707
6	preto	4	3	4,215
7	bois	2	2	4,135
8	brejo	2	2	4,135
9	mistérios	2	2	4,135
10	manhã	2	2	4,135
11	dali	2	2	4,135
12	céu	2	2	4,135
13	chegava	2	2	4,135
14	nuvens	2	2	4,135

Ordem	Léxico	Total	Hiato	Valor
15	nossas	2	2	4,135
16	onde	7	4	4,033
17	cavalos	7	4	4,033
18	sol	5	3	3,617
19	bobo	3	2	3,179
20	velho	6	3	3,162
21	o	254	42	3,156
22	bocejo	1	1	2,924
23	bojo	1	1	2,924
24	bem/bom	1	1	2,924
25	besta	1	1	2,924
26	bandos	1	1	2,924
27	bugresco	1	1	2,924

¹ Daqui por diante, utilizaremos o símbolo “§” como abreviação da palavra “parágrafo”, e simplificaremos a referência aos trechos citados de Guimarães Rosa, mencionando o parágrafo do trecho citado e o número da página correspondente.

A variável “Hiato” possui 21 vocábulos dentro da faixa privilegiada, entre os quais importam sobretudo os que carregam traços nocionais, como, por exemplo, “vaqueiro”, “manhã”, entre outros. Em relação ao *corpus* considerado, as duas colunas de valores de frequência – Total e “Hiato” – revelam os vocábulos exclusivos dessa variável, quando os valores são iguais, e aqueles que não são exclusivos da variável e, portanto, aparecem em outras variáveis do *corpus* selecionado, quando os valores divergem. A última coluna mostra o peso desses vocábulos, ou seja, a sua importância para a articulação do texto e do discurso, quando obtêm, mediante cálculos matemáticos realizados a partir de sua frequência e de suas relações com outros itens, o valor mínimo de 1,96.

Ainda dentro do valor preferencial acham-se os hápax, vocábulos com uma única ocorrência e exclusivos da variável em que aparecem. O fato de terem o peso acima do valor mínimo preferencial aponta sua importância na variável. Como a lista é extensa – são 328 hápax –, está exposto somente o começo da lista desses vocábulos, a partir do número de ordem 22, destacando-se a igualdade de valor entre as colunas Total e “Hiato” e a coluna com o peso significativo de 2,924. É preciso ressaltar que a lista preferencial não se esgota nessas duas menções, mas os vocábulos que aí aparecem são suficientes para qualquer procedimento voltado para o texto e para o discurso selecionados.

A TDVL possui outras faixas de vocábulos, reunidos segundo a sua importância: a zona intermediária corresponde ao *vocabulário básico*, subdividido em uma zona com tendência positiva e uma zona com tendência negativa. Trata-se do vocabulário de emprego corrente no sentido de formar o sustentáculo do discurso. Os essencialmente básicos têm o peso centrado em torno da média $-1 \leq Z \leq +1$, sendo indispensáveis à formação do texto e à delimitação específica da zona centrada no vocabulário comum. O vocabulário básico com tendência positiva, entre $+1 \leq Z \leq +2$, possui referenciais temáticos; o vocabulário básico com tendência negativa, entre $-2 \leq Z \leq -1$, contém mais itens responsáveis pela articulação e, grosso modo, sem referenciais temáticos. Situando-se além dos dois pontos negativos, o *léxico diferencial* é aquele cujo valor indica abandono ou rejeição, opondo-se ao preferencial enquanto carga de expressividade temático-discursiva. Dependendo da análise considerada, pode ser altamente significativo no *corpus* considerado. Nesta análise, será destacado o maior valor diferencial, o advérbio “não”, porque a rejeição no uso pode indicar uma tendência afirmativa do discurso.

Ordem	Léxico	Total	Hiato	Valor
573	ele	80	1	-2,694
574	não	155	3	-3,471

4. Descrição e análise da variável

4.1. Procedimentos gerais

Para a seleção de itens orientadores do sentido dentro da TDVL, é levada em conta a faixa essencialmente preferencial, com a identificação dos vocábulos de maior valor, podendo ser preferenciais, preferenciais-exclusivos ou hápax-preferenciais. Parte-se de itens lexicais situados entre os maiores valores e de suas relações com os demais itens lexicais preferenciais e/ou exclusivos, para estabelecer, por meio da análise de recortes discursivos, as relações que fundamentam a sintaxe e a semântica dos níveis profundo, narrativo e discursivo.

No que se refere à sintaxe do nível discursivo, a actorialização pede foco na flexão verbal de pessoa e na utilização dos pronomes pessoais e de tratamento. A falta desses identificadores na lista preferencial ou preferencial-exclusiva abre a perspectiva para a zona de articulação do texto, a zona de base, pois ela traz os fundamentos dessa articulação. A temporalização depende dos tempos verbais com maior valor e com maior incidência. A espacialização leva a referências espaciais figurativas como “bosque” ou “rancho”. Não havendo nenhum desses índices em evidência, identificam-se as debreações por meio da extração de recortes discursivos associados aos vocábulos de maior peso. É preciso lembrar que a actorialização, enquanto mecanismo da sintaxe discursiva, pode instalar a pessoa no enunciado por meio da debreação enunciativa da enunciação, o que pressupõe o espaço do *aqui* e o tempo do *agora*, ou por meio da debreação enunciativa da enunciação, constituindo o espaço do *lá* e o tempo do *então*. Esses indicadores podem ser ratificados pelos recortes discursivos da variável em análise.

O estrato básico e o diferencial são examinados naquilo que os ligam ao percurso gerativo de sentido, como os pronomes pessoais, identificadores da actorialização; o tempo das formas verbais, índices da temporalização, e as figurativizações, reveladoras do espaço.

Quanto às classes gramaticais, são considerados os vocábulos nocionais – substantivo, adjetivo, verbo e pronomes pessoais (reto, oblíquo e de tratamento) –, pertencentes à faixa preferencial ou preferencial-exclusiva. Nessa seleção, o vocábulo, enquanto item do léxico, é tratado como complexo narrativo em potencial, atualizado segundo a voz de quem o emprega, o contexto no qual é utilizado e as relações que mantém com os outros itens da construção discursiva. Geralmente, a lista de vocábulos preferenciais, ao identificar os itens de maior peso positivo, permite, por meio desse item, desenvolver todo o complexo descritivo e analítico de sentido, bastando identificar, nos recortes discursivos, a direção tomada. O léxico é a sede da

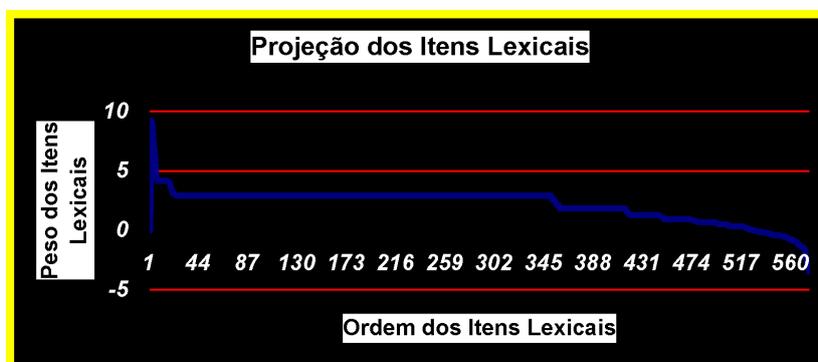
herança antropocultural do homem; logo, os itens lexicais contêm em latência os desdobramentos narrativos e discursivos, não se limitando à mera referência ou denominação. Nenhum item, considerado dentro do grupo nominal, é apenas referência; há nele uma representação multifacetada, cabendo ao enunciador ativar a rota pertinente aos seus objetivos de comunicação e ao co-enunciador distinguir o que foi estimulado.

Quanto às outras classes, como as palavras relacionais, os advérbios, os numerais, os artigos, os indefinidos, os demonstrativos e os possessivos, elas só serão consideradas quando estabelecerem, no discurso, uma relação fundamental com as categorias já destacadas.

4.2. Procedimentos específicos

Os procedimentos específicos dizem respeito às descrições das ocorrências relacionadas aos níveis do percurso gerativo de sentido. Na introdução de cada variável, as denominações utilizadas pertencem ao método Camlong; na descrição do percurso gerativo, a denominação *item lexical* é substituída por *figura*, quando pertinente. Seus pesos, dados pelo léxico preferencial, são mencionados, quando necessário. Os textos são tratados como *variáveis* durante toda a descrição e análise.

5. Descrição e análise da variável “Hiato”



A projeção de valores positivos no eixo y, o eixo vertical dos pesos, situa antes de dez o valor mais expressivo, identificando o vocábulo “Nhácio” com 9,246 na TDVL. O gráfico demonstra ainda a presença de itens entre essa marca e o ponto abaixo de 5, caracterizando a zona preferencial com a ocupação por itens lexicais que remetem ao homem em relação direta com a natureza: “vaqueiro” (7,162); “touro” (6,538); e, na pequena linha horizontal, “bois”, “brejo”, “manhã”, “céu” e “nuvens”, portadores do mesmo peso, 4,135. Abaixo desse valor, destacam-se “cavalos” (4,033) e “sol” (3,167).

Conforme foi mencionado no enfoque da TDVL, os 328 itens componentes dos hápax possuem o valor 2,924 e ocupam, pelo que demonstra a linha azul de projeção, a faixa horizontal mais extensa do gráfico acima, assegurando, de imediato, um rol lexical extenso e singular, que influencia significativamente o desdobramento do discurso. Na faixa de rejeição, a faixa negativa do gráfico, a linha azul atinge a zona do peso -3,471, ratificando os valores da TDVL, pela qual se identifica o vocábulo “não” como portador desse valor, o que sinaliza uma tendência afirmativa presente no discurso da variável em análise.

5.1. Nível fundamental

Entre os itens lexicais que atendem ao critério de seleção prévia, isto é, são preferenciais, destacam-se “manhã”, “sol” e “preto”. Eles sinalizam possíveis articulações do nível fundamental do percurso gerativo de sentido, à medida que mantêm relação entre si, indicando matizes contrários da percepção visual. Os termos trazem como referência as oposições luminosidade e escuridão, claro e escuro, luz e sombra. Procedendo-se às extrações, a fim de ratificar as ocorrências no texto, obtêm-se os seguintes resultados:

- Manhã e sol
 - Luz/luminosidade, claro/claridade em relação com a natureza:

Iam os cavalos a mais — o céu sol, massas de luz, nuvens drapuxadas orvalho perla a pérola (p. 72, § 2).

A manhã era indiscutível (p. 72, § 2).
 - Luz relacionada à vida:

A manhã por si respirava (p. 72, § 5).

A expressão “Ia tudo pelo claro” (p. 72, § 5) condensa as observações anteriores, demonstrando a relação entre luz, vida e natureza, e acrescentando semas metafóricas que fazem inferir o reconhecimento do habitual, a ausência de qualquer perturbação.

- Preto

- A comparação da cor do touro com o pano preto estabelece uma relação da natureza com o conhecimento construído pela cultura. Antes de identificar o ser, há a referência ao “preto” do tecido, o que coloca a observação da cor em primeiro plano, no que diz respeito ao touro:

Preto como grosso esticado pano preto, crepe, que e quê espantoso (p. 73, § 6).

- Associação entre a ancestralidade, o medo, o touro e a cor preta, cujo simbolismo constitui um saber cristalizado, de origem desconhecida:

“Remoto, o touro, de imaginação medonha – a quadratura da besta – ingenerado preto empedernido” (p. 73, § 10).

As observações a seguir reforçam a relação do preto e seus elementos isotópicos – escuro, trevas, sombra – com o desconhecimento e a morte. O “touro” escuro surge como representação do mal, da morte, no conhecimento acumulado: “Errático, a retrotempo, recordava-se sobre nós o touro, escuro como o futuro, mau objeto para a memória” (p. 73, § 12).

Com essas extrações, emergem a luz, a manhã e a natureza como elementos positivos, representativos da *vida* (“A manhã por si só respirava”, p. 72, § 5), o que constitui de imediato seus contrários, o preto, as sombras e a cultura, como pólo negativo, relacionado à *morte*. A foria aí presente, ou seja, a escolha dos valores eufóricos e disfóricos, respectivamente valores tidos como positivos e valores percebidos negativamente, que já se manifesta no nível fundamental (Tatit, 2001, p. 17-20), reforça-se nas palavras dos locutores Nhácio e Põe-Põe, no parágrafo 3, e na descrição do touro por parte do vaqueiro, associando os chifres à foice, um símbolo da morte:

- Nhácio:

“Iii, xem, o bem-bom, ver a vez de galoppear [...]”

- Põe-Põe:

“Ih, é, ah! Ô vida para se viver!” (p. 72, § 3). Touro mor que nenhuns outros, e impossível, nuca e tronco, chifres feito foices, o bojo,

arcabouço, desmesura de esqueleto, total desforma (p. 73, § 6).

Os conteúdos mínimos a impulsionar o sentido são, então, *vida* e *morte*, o primeiro é eufórico e o segundo, disfórico. A natureza concretiza a *vida* plena, pelo modo como é descrita: muitos dos hápax são referências acuradas da flora do espaço local, onde, felizes, os homens cumprem as suas tarefas. O antropônimo palatalizado “Nhácio” carrega o maior valor da *TDVL*, surgindo na narrativa como herói modelar do espaço (§ 4). Entretanto, no relato, ao conhecimento herdado corresponde muitas vezes o desacerto de uma cultura que nada assegura ou prevê: o touro não desenvolve o comportamento esperado, como observam Nhácio, no parágrafo 16, e o narrador, em seus comentários. A singularidade do acontecimento inscreve a foria presente no nível fundamental:

- Euforia – *vida*

“Iam os cavalos a mais – o céu sol, massas de luz, nuvens drapuxadas orvalho perla a pérola. Refartávamos de alegria e farnel” (p. 72, § 1).

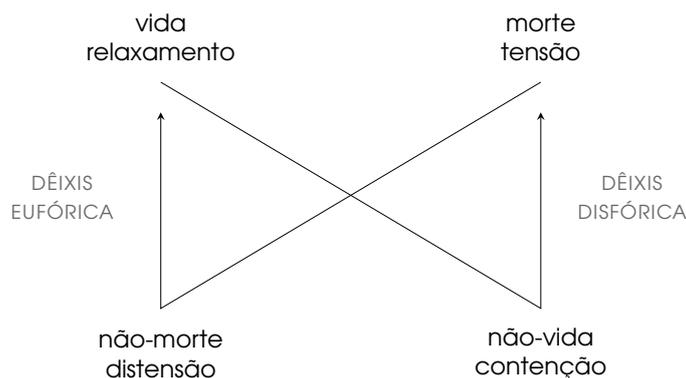
- Disforia – *morte*

Touro mor que nenhuns outros, e impossível, nuca e tronco, chifres feito foices, o bojo, arcabouço, desmesura de esqueleto, total desforma. Seu focinho estremeceu em nós, hausto mineral, um seco bulir de ventas – sentíamos sob as coxas o sólido susto dos cavalos. Olhos – sombrio e brilho – os ocos da máscara. Velho como o ser, odiador de almas (p. 73, § 6).

Remoto o touro de imaginação medonha – a quadratura da besta – ingenerado, preto, empedernido. Ordem de mistérios sem contorno em mistérios sem conteúdo (p. 73, § 10).

Mesmo nem nos maleficiara – com nenhum agouro, sorvo de sinistro – o estúrdio bronco monstro. De onde vem então o medo? (p. 73, § 17).

As valorizações presentes no nível fundamental não representam somente um percurso a ser revestido pelos outros níveis, elas revelam as tendências de um enunciador, elas identificam aspectos do *éthos*. Aqui, a *vida*, refletida na natureza exuberante, como descreve o vaqueiro narrador, sobrepõe-se à *morte*, relacionada ao que a cultura acumula e transmite como herança às sucessões.



Há um percurso, partindo do termo *vida*, com o estado relaxado, o “estar à vontade” dos vaqueiros, que, repentinamente, é interrompido pela figura do touro, trazendo a contenção, a não-disforia e, por fim, a tensão de quem tem diante de si o elemento disfórico que é a *morte*, representada pelo touro preto. Este, quando se comporta em desacordo com o esperado, provoca a distensão da *não-morte*, que, por fim, desemboca na *vida* encontrada no rancho: o repouso, o alimento, a troca, itens essenciais a todos os seres.

5.2. Nível narrativo

Se a sintaxe do nível narrativo distingue enunciados de estado, em que o sujeito se acha em conjunção ou em disjunção com o objeto valor, e enunciados de fazer, com um programa de liquidação da falta, pode-se dizer que a narrativa em questão aponta para a plenitude: os sujeitos estão conjuntos à natureza, à *vida*, que é valorizada positivamente nas categorias de base.

Há, no entanto, um conhecimento prévio, persuasivo, refletido na simbologia que o touro preto carrega. O saber e o crer estimulam a percepção da morte e a intimidação, propondo ao manipulado, representado pelo narrador e também pelo grupo no qual se insere, uma sanção negativa prévia, aplicada ao ser enorme e escuro:

Remoto, o touro, de imaginação medonha — a quadratura da besta — ingenerado, preto, empedernido (p. 73, § 10).

Errático, a retrotempo, recordava-se sobre nós o touro, escuro como o futuro, mau objeto para a memória (p. 74, § 12).

Empatara-nos, aquele, em indisarce, advindamente; perseguia-nos ainda, imóvel, por pavores, no desamparadeiro (p. 74, § 14).

Mesmo nem nos maleficiara — com nenhum agouro sorvo de sinistro — o estúrdio bronco monstro (p. 74, § 17).

Os valores atribuídos à forma e à cor, a princípio, intimidam. Entretanto, a série de acontecimentos previstos não se concretiza (p. 73, § 6); o elemento gerador do medo que imobiliza o grupo e também assusta os

cavalos não desenvolve o percurso esperado: o ataque, a luta, a morte.

Era enorme e nada (p. 73, § 6).

Algum turbar entrecontagiava-nos, sem reflexa o útil (p. 73, § 8).

Remoto, o touro, de imaginação medonha — a quadratura da besta — ingenerado, preto, empedernido. Ordem de mistérios sem contorno em mistérios sem conteúdo. O que o azul nem é do céu: é de além dele. Tudo era possível não acontecido (p. 73, § 10).

De onde vem então o medo? Ou este terraqueo mundo é de trevas, o que resta do sol tentando iludir-nos do contrário (p. 74, § 17).

O sujeito vaqueiro, antes pleno com relação a ser e saber, conhecedor dos companheiros, dos caminhos e dos animais, vê-se tomado pelas emoções, pelo medo, pelas pulsões. Enquanto sujeito destinatário de um saber, é levado a refletir sobre o conhecimento que traz, em que não há lugar para o imprevisível: a morte que desponta; a morte que não acontece. Como testemunha do oculto que se inscreve no óbvio (“O redor o olhava”, p. 73, § 9), adquire uma outra percepção de si e do grupo. As sombras a que se refere em “Fazia cansaço no furto frio de nossas sombras” (p. 73, § 17) tornam-se o desconhecimento coletivo, o medo e ainda uma referência ao passado e às origens.

5.3. Nível discursivo

5.3.1. Actorialização

Na zona preferencial, ao lado de “sol”, “manhã” e “preto”, destaca-se a forma preferencial “chegava”, associada ao pronome “se”, indicador de indeterminação, que, na narração, inicia e encerra o relato:

Redeando rápido, como jovem vaqueiro Põe-Põe e o vaqueiro velho Nhácio, chegava-se à Cambaúba [...] (p. 72, § 1). Ainda, pois, chegava-se — ao rincão, pouso, tetos — rancharia de todos (p. 75, § 24).

A lista de hápax aponta outras realizações de igual valor no que diz respeito à indeterminação:

Avançava-se — a torto avançava-se (p. 73, § 7). Havia-se — subindo ainda às nuvens de onde havia-se de cair (p. 72, § 5). Vinha-se — Vinha-se levíssimo, nos animais (p. 72, § 5).

No entanto, ainda na zona preferencial, a ocorrência de nossas mostra a primeira pessoa, o que também se verifica nas formas verbais inseridas na mesma faixa de destaque lexical. Evidenciam-se, assim, um narrador actante incluído em um grupo e participante do percurso que descreve:

Fazia cansaço, no furto frio de nossas sombras. Tirávamos passo (p. 74, § 17). Refartávamos de alegria e farnel (p. 72, § 2). [...] sentíamos sob as coxas o sólido susto dos cavalos (p. 73, § 6). Mas montávamos à área da colina (p. 73, § 11). Topávamos rede, fogueiro, prosa, paz de botequim, à qualquer conta (p. 75, § 24).

A oscilação entre duas projeções — a indeterminação e a primeira pessoa do plural — delineia um narrador que procura evitar o realçamento da subjetividade, colocando-se ora como observador imparcial, ora como participante de um grupo, portador das mesmas crenças e sujeito aos mesmos desígnios.

Fiorin (2002, p. 118) alerta para a existência de procedimentos macrotextuais. No texto analisado, verifica-se a presença de emblemas em que o narrador escolhe um foco que orienta, na verdade, para outro. Assim, este narrador, quando opta pela indeterminação, na verdade fala de si mesmo, do grupo do qual faz parte e ainda do narratário, traçando um efeito de neutralidade, que pode ser identificado na expressão passiva “se vê” (p. 72, § 1). O que aborda vale indistintamente para todos e para ele mesmo. Quando se expressa como *nós*, enfatiza a intensidade de sua ligação com o que expressa: os sujeitos, incluindo a si próprio, acham-se sob as mesmas determinações. A forma *nós* indica a ligação com os demais, entre os quais se encontra o narratário, e marca o ponto de observação, que dá margem às reflexões pessoais (“Mesmo nem nos maleficiara — com nenhum agouro, sorvo de sinistro — o estúrdio bronco monstro”, p. 74, § 17).

Esse narrador cede a voz a dois interlocutores, os companheiros de jornada *Põe-Põe* e *Nhácio* (p. 72-75, §3, 9, 16, 19, 20, 22 e 23), fazendo prevalecer a fala destes, que, algumas vezes, ratifica suas considerações:

Narrador: Refartávamos de alegria e farnel. A manhã era indiscutível. Tantas vias e retas.

Nhácio: “Iii, xem, o bem-bom, ver a vez de galopear [...]”

Põe-Põe: “Ih, é, ah! Ô vida para se viver!” (p. 72, § 1 e 2).

Narrador: Algum turbar entrecontagiava-nos.

Nhácio: “Ixe, coragem também carece de ter prática!” (p. 73, § 8 e 9).

Algumas vezes, as formulações respondem a indagações do narrador que projetam a *performance* do narratário; outras vezes, antecipam as informações do narrador ou estimulam as reflexões:

Narrador/narratário: O touro?

Nhácio: “Mas, é um marruás manso, mole, de vintém! Vê que viu a gente, encostados nele, e esbarrou, só assustado, bobo, bobo”

Narrador: “Mesmo nem nos maleficiara — com nenhum agouro, sorvo de sinistro — o estúrdio bronco monstro” (p. 74, § 15, 16 e 17).

A actorialização revela, portanto, um narrador actante, figurativizado como vaqueiro, que procura diluir-se no grupo, porque submetido às mesmas determinações. As indagações utilizadas inserem o narratário naquilo que têm como foco. É preciso ressaltar que o narrador não se diz vaqueiro. Há uma elipse da informação, que pode ser inferida de outras figuras como “redeando”, “escolta”, “campeação”; da profissão dos companheiros de jornada e da referência aos “bois” no parágrafo 13.

5.3.2. Temporalização

O marco temporal é dado pelo emprego do presente do indicativo com aspecto atemporal, à medida que caracteriza as referências descritivas feitas à natureza. Para focalizar um dia de trabalho no tempo não-concomitante ao momento de seu relato, o narrador e ator vaqueiro opta pelo pretérito imperfeito, caracterizando a debreagem temporal enunciativa. O uso do pretérito imperfeito revela uma visão interna dos acontecimentos, interrompida por ações pontuais, expressas no pretérito perfeito 2:

Redeando rápido, com o jovem vaqueiro Põe-Põe e o vaqueiro velho Nhácio, chegava-se à Cambaúba, que é um córrego, pastos, onde se vê voam o sai-xê [...] e canta o ano todo a patativa (p. 72, § 1).

Era enorme e nada. Reembrenhou-se (p. 73, § 6).

Na articulação entre tempo concomitante e tempo não-concomitante ao momento do relato, inserem-se reflexões e apreciações marcadas pelo aspecto atemporal, seja pela forma verbal no presente do indicativo, seja pela ausência de verbos. Essas ocorrências levam à instância da enunciação e, conseqüentemente, à

presença do enunciador e, também, de seu par complementar, o co-enunciador, especialmente quando é utilizada a indagação (p. 72-74, § 5, 10 e 17).

No parágrafo 10, as referências a um tempo remoto estão presentes nas reflexões sobre natureza e cultura, revelando a preocupação com as origens. Para o enunciador, nas sucessões da vida, o imprevisível e o inexplicável sobrepõem-se ao saber partilhado, da razão, do conhecimento, apontando, assim, para a impotência dos seres. Utilizando a indeterminação e a primeira pessoa do plural no processo de actorialização, demonstra a sua inclusão nesses mesmos desígnios.

5.3.3. Espacialização

A forma verbal “chegava”, entre os preferenciais, leva às referências “Cambaúba” e “rancho”, do primeiro e do último parágrafos, identificando espaços e movimento. Quando tal forma verbal aparece relacionada a verbos flexionados no presente — “é”, “vê”, “canta” (p. 72, § 72) —, revela-se a oscilação entre a debreagem enunciativa e a debreagem enunciativa projetadas na configuração espacial. A primeira traz o local dos acontecimentos narrados; a segunda caracteriza a perspectiva que engloba o relato, reforçada pelo advérbio “lá” no quinto parágrafo.

O lugar enunciativo pode ser circunscrito pela expressão “Desta maneira” (p. 75, § 24). ela permite inferir a passagem do ambiente em que aconteceram o descanso e a prosa no pretérito para o tempo presente, instalando o contexto do “rancho” no espaço da enunciação, que caracteriza o ambiente e as relações entre enunciador e co-enunciador. O espaço enuncivo, o espaço dos acontecimentos narrados, é marcado pelo movimento dos actantes-atores, expresso em referências como “após”; “Vinha-se levíssimo”, “subindo”, “havia-se de cair.”; “quebravam os cavalos à mão direita”; “Tinha-se, sem querer, dado rodeio, tirando do caminho afastamento de grande arco”; “adiante”; “chegava-se” (p. 72-75). Os trechos citados permitem depreender a categoria horizontalidade no *vs* verticalidade, ratificada sobretudo na expressão “afastamento de grande arco” (p. 73, § 10). A expansão dentro do espaço natural é expressa na marcha interrompida pela verticalidade, manifestada em “à beira de pirambeira” (p. 73, § 9); o risco da altitude associa-se à consciência do imponderável experimentado pelos actantes no encontro com o touro. O termo “colinas” (p. 74, § 11) reflete um princípio de distensão após a tensão gerada pela experiência do imprevisível, distensão essa que se reforça no parágrafo 13 e complementa-se no espaço do “rancho”, citado no último parágrafo.

5.3.4. Figuras e temas

Todas as indicações feitas pela faixa de peso mais expressiva da TDVL impulsionam ou constituem os percursos figurativos e temáticos: “manhã”, “claro”, “luz”

de um lado; “touro preto”, “sombras”, “escuro”, de outro. “Chegava” e “vaqueiro” trazem os caminhos e os animais. A reunião de todos esses grupos constrói a oposição em que se enaltece a vida por meio da descrição positiva da natureza e se rejeita a morte, cuja imprevisibilidade coloca em relevo o desconhecimento e o medo. A alternância entre debreagens e embreagens de pessoa reúne narrador e narratário, faz emergir o enunciador e traz o co-enunciador para as mesmas determinações da *vida* e da *morte*. Construindo uma cena de prosa e camaradagem, que pode ser identificada no último parágrafo, em especial na expressão “Desta maneira” (Rosa, 1985, p. 75), o enunciador focaliza uma jornada cotidiana em meio à natureza exuberante e poderosa, em que tudo caminha por si (primeiro e segundo parágrafos). Alvo de uma atitude descritiva, o espaço se reveste de uma figurativização do mundo natural — vegetal, animal e cósmico —, ratificada pela isotopia composta por “animais”, “árvores”, “bosques”, “buritis”, “caparrosa”, “cerrado”, “chifres”, “colinas”, “córrego”, “ervas”, “focinho”, “folhagens”, “gaia”, “galhos”, “grotas”, “lagoa”, “maria-branca” ou “maria-poesia”, “marrecas”, “matas”, “papagaios”, “pássaros”, “pastos”, “patas”, “patativas”, “pedras”, “plantas”, “pirambeira”, “ramos”, “saí-xê”, “sementes”, “tronco”, “vargedos”, “vegetais”, “ventos”, “xexéu”. Todas as referências pertencem aos hápax que se acham na zona preferencial. Com um vocabulário particular e preciso, enfatiza-se uma natureza exuberante e enaltece-se a vida, o que reforça a valorização das categorias de base do nível fundamental.

Há figuras ligadas ao reino animal — “cavalo”, “bois”, “touro” — e ao mundo natural de caráter cósmico, pois tanto ocorre uma referência à terra — “brejo” —, como ao próprio “céu” e às “nuvens”. Constrói-se uma verticalidade revestida pela amplitude e matizada pelo termo “manhã”, que instala o valor da claridade, da luz, na cena que se forma.

O elemento humano comparece integrado a esse entorno não só por meio da indicação profissional — “vaqueiro” — que se associa aos elementos destacados do reino animal — “bois” e “cavalos” —, mas também porque homem e animais se igualam na medida em que estão sujeitos às mesmas determinações: os vaqueiros e os cavalos reagem com o mesmo temor diante da figura ameaçadora do touro, como relata o narrador no parágrafo 6.

O cromatismo é destacado por meio da presença da luminosidade, da cor preta e da observação do entorno, ratificando a percepção visual como predominante. A trajetória da ação no espaço marcado pela amplitude e pela diversidade dos reinos indica o corpo ativo, ágil e vigoroso de um homem também observador e reflexivo, que observa os caminhos, os companheiros, os acontecimentos e emite seu julgamento. Isso se torna perceptível à medida que formas e cores são preponde-

rantes para marcar as valorizações (p. 72 e 74, § 1, 2, 5, 17).

A figura do homem-vaqueiro na lida rotineira, em caminhos que identifica e descreve com propriedade — o que explica o alto valor dos hápax —, reflete a crença no conhecimento que estabelece controle e direção da vida, em que não cabe o imponderável. Os vaqueiros parecem saber de seu entorno, dirigem os animais e cumprem o seu dia exatamente como em qualquer outro cotidiano de qualquer parte. As “Tantas vias e retas” do enunciador (p. 72, § 2) advertem sobre o imprevisível: a vida não é só o que se apresenta. Enfatizando a percepção visual na descrição da luminosidade, busca chamar a atenção para o oculto, que surge repentinamente na figura ostensiva de um touro preto. Em meio à valorização da natureza, faz uma contínua referência à ancestralidade, ao remoto, insere as pulsões por meio do medo e define suas origens como mistério, o que possibilita deduzir a insuficiência do saber acumulado pela civilização. O tom ameno, de camaradagem, revela o *éthos* de um sujeito que, tal como o vaqueiro, procura ser o guia que envolve e desperta a atenção para as trilhas insondáveis e sem luz (“Fazia cansaço no furto frio de nossas sombras”, (p. 74, § 17). Com elas, emerge a impotência dos sujeitos, a impossibilidade dos seres cujo saber e cuja experiência não são suficientes para ver, prever, descobrir e assegurar.

Considerações finais

Os dados fornecidos pela Tabela de Discriminação de Valores Lexicais mostram-se produtivos: destacam as principais figuras, e estas, por sua vez, geram as relações nos recortes discursivos, instalam e constituem as rotas essenciais do percurso gerativo de sentido e as consequentes observações. Os valores pertencentes à zona preferencial colocam em evidência figuras como “vaqueiro”, “touro”, “cavalos”, “bois”, “céu”, “manhã”, “preto”, “mistérios”, “nuvens”, “sol”, “nossas”, que formam o complexo de toda a articulação descrita, fundamentada em *luz* e *sombra*. O alto valor dos hápax sinaliza a descrição precisa do entorno por parte do vaqueiro, a valorização da natureza e da vida e as reflexões sobre o conhecimento, o imprevisível e o medo.

Há uma atitude afirmativa nas testificações relacionadas à natureza, nas reflexões que o circundante desperta e nas lições que o imprevisível ministra, já sinalizada por oposição ao valor de rejeição do advérbio “não” na TDVL, tanto no que se refere às cenas que possuem o aspecto atemporal, quanto às do pretérito:

[...] que é um córrego, pastos, onde se vão o saí-xê, o xexéu [...] e canta o ano todo a patativa (p. 72, § 1).

[...] bosques da caparrosa comum surpre-

endem [...] Tantas vias e retas (p. 72, § 2).

Toda água é antediluviana. [...] A manhã, por si, respirava. Macegão: lá o angola cresce, excele [...] (p. 72, § 5).

Velho como o ser, odiador de almas [...]. De temor, o cavalo ressona, ronca, uma bulha nas narinas, como homem que dorme (p. 73, § 6).

Remoto, o touro, de imaginação medonha — a quadratura da besta — ingenerado, preto empedernido. Ordem de mistérios sem contorno em mistério sem conteúdo. O que o azul nem é do céu: é de além dele (p. 73, § 10).

[...] escuro como o futuro, mau objeto para a memória (p. 74, § 12).

Era, sim, casado, o vaqueiro Nhácio, carafuz (p. 74, § 18).

[...] rancharia de todos [...]. Desta maneira (p. 75, § 24).

Em paralelo à atitude reflexiva, que revela suas observações e suas constatações, delinea-se um tom ameno, de camaradagem (p. 75, § 24), a manifestar um *éthos* de um sujeito que, tal como o vaqueiro, procura ser o guia e o camarada que envolve e desperta a atenção para as trilhas insondáveis e sem luz (“Fazia cansaço, no furto frio de nossas sombras”, p. 74, § 17), das quais emerge a impotência dos sujeitos. Nessa perspectiva, o título “Hiato” reflete as intenções presentes no discurso, a referência ao vazio do conhecimento. Ao mesmo tempo, aplica-se ao aparecimento do touro e à consequente interrupção de uma rotina alegre, bem como à pausa final para a alimentação, o descanso e a prosa conscientizadora (p. 75, § 24).

As observações feitas acima, bem como as que foram desenvolvidas ao longo do trabalho nascem com os itens privilegiados pelos valores da TDVL, produzida pelo método Camlong: o título “Hiato” é hápax preferencial, com o peso de 2,924 na variável. Os demais itens em evidência em toda a descrição, inseridos no percurso gerativo, transformados em figuras relacionadas pela articulação discursiva, igualmente delinham rotas fundamentais, constituem as isotopias e os temas dominantes, e pontuam os comentários. O diálogo entre a tecnologia computacional, a estatística e a semiótica de textos de linha francesa fica estabelecido. ●

Referências

- Barros, Diana Luz Pessoa de
1988. *Fundamentos semióticos*. São Paulo: Atual.
- Barros, Diana Luz Pessoa de
1990. *Teoria semiótica do texto*. São Paulo: Ática.

- Camlong, André
1996. *Methode d'analyse lexicale, textuelle et discursive*. Paris: Ophrys.
- Durand, Gilbert
1989. *Estruturas antropológicas do imaginário*. Lisboa: Presença.
- Fiorin, José Luiz
1989. *Elementos de análise do discurso*. São Paulo: Contexto.
- Fiorin, José Luiz
2002. *As astúcias da enunciação*. São Paulo: Ática.
- Greimas, Algirdas Julien; Courtés, Joseph.
[1983]. *Dicionário de semiótica*. São Paulo: Cultrix.
- Rosa, João Guimarães
1985. *Tutaméia: terceiras estórias*. Rio de Janeiro: Nova Fronteira.
- Tatit, Luiz
2001. *Análise semiótica através das letras*. São Paulo: Ateliê Editorial.
- Zapparoli, Zilda; Camlong, André
2002. *Do léxico ao discurso pela informática*. São Paulo: Edusp/Fapesp.

Dados para indexação em língua estrangeira

Santos, Márcia Angélica dos
Dialogue between Technology, Statistics and Semiotic to an Approach to Texts
Estudos Semióticos, vol. 5, n. 2 (2009)
ISSN 1980-4016

Abstract: *The approach developed in this paper aims to verify the possibilities of qualitative and quantitative contents produced by the Table of Discrimination of Lexical Values as a database to the development of other approaches. This table is part of André Camlog's instrumental methodology (1996), from University of Toulouse II, which is exclusively related to description and analysis of lexicons, texts, and discourses. Its information is produced by mathematical and statistic operations, processed through computer technology. By transforming texts into measured lists of words, it makes this information available to other approaches, independent of the process related to the approach. To investigate the length of this data application, one must consider the Table of Discrimination of Lexical Values as a resource of the Meaning Generative Process proposed by Greimas and Courtés (1983), and its levels: basic, narrative and discursive. Taking into account syncretic enunciation – on the one hand statistic data produced by technology; on the other hand, generative process in its more traditional form, this research concludes that lexical items, organized according to frequency and value related to its environment, are enough to support the semiotic approach to textual analysis, establishing safety of scientific description and validating the development of any methodology.*

Keywords: *computer technology, statistics, lexicons, semiotics*

Como citar este artigo

Santos, Márcia Angélica dos. Diálogos entre a tecnologia, a estatística e a semiótica para uma abordagem de textos. *Estudos Semióticos*. [on-line] Disponível em: { <http://www.fflch.usp.br/dl/semiotica/es> }. Editores Responsáveis: Francisco E. S. Merçon e Mariana Luz P. de Barros. Volume 5, Número 2, São Paulo, novembro de 2009, p. 98-108. Acesso em “dia/mês/ano”.

Data de recebimento do artigo: 09/12/2008

Data de sua aprovação: 26/02/2009
