

# Gêneros Digitais: Os *Search Engines* Como Interlocutores

## No Diálogo Em Meio Digital

Marco Antônio Gutierrez<sup>1</sup>

**Resumo:** Apesar da consciência crescente de que o ciberespaço implica práticas discursivas novas (FAIRCLOUGH, 1996), as pesquisas disponíveis estão longe de firmar consenso sobre o alcance dessa revolução nos modos de interação social. Ressalta aqui o lugar especial ocupado pelo problema dos gêneros digitais, cuja abordagem preliminar parece indispensável para caracterizar esse novo campo da atividade discursiva. Nesse território inexplorado, é natural a inexistência de consenso sobre a caracterização de inúmeros eventos comunicativos (SWALES, 1990), em particular os produzidos por dispositivos computacionais como os *search engines*. Esta pesquisa parte da teoria dos gêneros esboçada por Bakhtin e procura determinar o lugar das ferramentas de busca entre as práticas discursivas do ciberespaço. Através de buscas experimentais, procuramos determinar como as diferentes concepções da audiência programadas nas heurísticas dos dispositivos afetam o resultado das pesquisas, caracterizando um diálogo onde o dispositivo computacional exerce papel ativo de interlocutor. O produto intertextual de tais buscas é marcado mais pela encenação do discurso do outro através do hipertexto que pela sua representação e parece indicar que diferentes dispositivos respondem de modo diferente às mesmas buscas, por força de escolhas baseadas numa concepção distinta da audiência, caracterizando uma *vontade discursiva* (Bakhtin, 2003) própria.

O ciberespaço é um território social avesso ao controle, onde a liberdade pode se expressar plenamente, para o bem ou para o mal. Essa tendência é sustentada pelas próprias características tecnológicas da Internet, em especial pelo fato de se tratar de uma rede sem centro e virtualmente ilimitada, o que inviabiliza a adoção de mecanismos de controle eficazes. No entanto, a ordem do discurso de que falava Foucault (2004) também se infiltra no ciberespaço, exprimindo-se numa contra-tendência centrada nas tentativas de controle da circulação irrestrita da informação. Evidentemente, esse esforço de controle do discurso tem inúmeras vertentes com graus variados de eficácia. Uma das mais eficazes pela sua viabilidade consiste em exercer algum tipo de controle nesses centros de informação que são os dispositivos de busca.

O controle exercido nos dispositivos de busca é um mecanismo de controle do discurso que procura censurar a enunciação por meios indiretos, controlando a palavra do outro: ao restringir o acesso a documentos que incorporam temas “proibidos”, o que governos como o chinês, por exemplo, querem restringir a vigência no discurso de múltiplas vozes, perigosas porque diferentes, substituindo-as pelo discurso monótono, obliterando-as para o território do não-dito (porque não-lido). É essa uniformidade de um discurso plano o que se

---

<sup>1</sup> Orientadora: Anna Elizabeth Balocco

quer alcançar com a censura. No entanto, esses mecanismos de controle externo tentados por alguns governos sempre se revelaram de curto alcance em longo prazo, sendo, portanto, ineficazes. Muito mais eficaz que a censura é a auto-censura, sobretudo quando assumida como expressão da liberdade para a palavra do outro...

Neste trabalho abordaremos o funcionamento de um desses mecanismos de auto-censura incorporados aos dispositivos de busca, assumindo que eles têm como efeito não apenas o controle do próprio discurso, como também o do discurso do outro – o usuário que busca – ao delimitar (e, implicitamente, limitar) o universo de discursos a que temos acesso no ciberespaço. Os dispositivos de busca desempenham um papel essencial no acesso à informação na Internet, papel indispensável desempenhado na prática por apenas três buscadores – o Google, o Yahoo e o MSN, com 49,2%, 23,8% e 9,6% das 213 milhões de buscas diárias, respectivamente (ver SULLIVAN, 2006). Portanto, do ponto de vista prático, o acesso à informação na Internet é controlado por esses três dispositivos. O que sustentamos é que tais dispositivos controlam não apenas o acesso à informação, mas também, em grande medida, sua própria produção. Isso decorre da concepção de discurso com a qual privilegiadamente trabalhamos – a de Bakhtin.

Se “todo enunciado concreto é um elo na cadeia da comunicação discursiva de um determinado campo” (BAKHTIN, 2003: 296), meu enunciado de hoje é também uma resposta a um conjunto de enunciados de outrem, com quem dialoguei ontem. Ora, ao controlar em grande medida as vozes com quem sou capaz de dialogar na Internet, os dispositivos de busca, em longo prazo, são capazes de controlar também, em alguma medida, o meu próprio discurso. Se o ciberespaço fosse o único espaço social de interação discursiva disponível, os resultados das buscas nos *search engines* abarcariam praticamente todos os enunciados com os quais sou capaz de dialogar. Cremos que essa é uma razão muito boa para tornar essencial o estudo dos dispositivos de busca do ponto de vista discursivo.

Qualquer um que já tenha utilizado um *search engine* saberá que o crítico nesses dispositivos não é o acervo – geralmente envolvendo bilhões de documentos – mas o critério de ordenação e apresentação dos resultados, isto é, a classificação da relevância das saídas. Uma simples busca pode envolver milhares de documentos como saída – e obviamente ninguém é capaz de inspecionar todos. Em pesquisa recente, a iProspect (IPROSPECT, 2006) concluiu que 88% dos usuários abandonam uma busca após consultar apenas a terceira página de resultados; 41% dos usuários abandonam a busca já após a primeira página!

Isso implica que o conteúdo do acervo dos dispositivos é irrelevante para o resultado concreto da interação de busca porque a quantidade de documentos coerente com a expressão

buscada é sempre muito maior do que a nossa capacidade de inspecioná-los. Desse modo, o fator mais relevante na busca deve ser o critério de classificação do acervo e de ordenação das saídas para apresentação. É aqui – na determinação do que se convencionou chamar o grau de *relevância* de um documento para a busca que o usuário está realizando – que os projetistas de sistemas devem fazer algumas suposições críticas sobre o que os usuários querem obter quando interrogam o dispositivo em suas buscas no ciberespaço.

A idéia na origem dos dispositivos de busca é coletar, armazenar e eventualmente localizar o que se convencionou chamar *documentos relevantes*, isto é, aqueles documentos disponíveis na Internet nos quais o usuário da busca está interessado. Essa é uma tarefa impraticável, ainda que teoricamente possível: a quantidade de documentos disponíveis na Internet é muito grande e cresce continuamente. Além disso, uma fração não desprezível desses documentos é continuamente removida e atualizada. Esses fatores aliados à limitada capacidade computacional dos dispositivos fazem com que o conjunto de respostas de uma busca deva ser diferente em algum grau daquele conjunto (teórico) de documentos relevantes.

Seja  $R$  o conjunto de documentos relevantes no universo da Internet e  $|R|$  o número de elementos em  $R$ ; seja  $D$  o conjunto de documentos obtidos por um dispositivo de busca e  $|D|$  o número de elementos em  $D$ . O alvo do processo de busca deve ser, então, obter o conjunto intersecção de ambos, isto é:  $R_d = R \cap D$ . Chama-se cobertura (*recall*) a fração  $C$  de documentos relevantes recuperados na busca, isto é, a razão entre  $|R_d|$  e  $|R|$ ; chama-se precisão (*precision*) a fração  $P$  de documentos recuperados que é relevante, isto é, a razão entre  $|R_d|$  e  $|D|$ . Como não é praticável obter  $C = 1$ , o objetivo dos projetistas desses dispositivos é aumentar o valor da precisão  $P$  para valores tão próximos da unidade quanto possível. Duas estratégias são aqui utilizadas. A primeira delas é de natureza estritamente tecnológica: a indexação dos documentos deve ser tal que assegure ao máximo possível que os diferentes assuntos abordados num documento que eventualmente possam ser relevantes numa busca estejam acessíveis durante o processo. Como resultado, os dispositivos refinam continuamente sua capacidade de indexação, não apenas de modo a abarcar todo o conteúdo de um documento, independente do seu formato, como também de modo a catalogar elementos do “contexto” do documento, incluindo coisas como o diretório virtual onde reside, o nome do documento, documentos vizinhos, etc. A maior parte das pesquisas nesse domínio acadêmico visa elevar e acelerar essa capacidade de indexação.

Uma outra estratégia, porém, é também adotada, embora tenhamos encontrado notavelmente poucos estudos sobre o assunto: trata-se de elevar a capacidade do dispositivo de determinar o que o usuário está buscando. Essa estratégia permite que os dispositivos

selecionem do acervo de indexação somente aquelas entradas relevantes para o usuário. Quanto maior a capacidade de o dispositivo interpretar essa intenção do usuário, mais preciso será o resultado de  $P$ . Mesmo isso pode ainda resultar num valor muito elevado para  $D$ , mesmo que a razão  $R_d$  seja ótima (igual à unidade). Desse modo, os dispositivos devem igualmente classificar o grau de relevância de um documento e ordenar sua saída de acordo.

Suponhamos que o documento  $a$  qualquer aborde principalmente o tema  $x$  e marginalmente o tema  $y$ , enquanto que o documento  $b$  faça o inverso. Esse conhecimento deve ser expresso de algum modo no dispositivo, de tal modo que um usuário que esteja pesquisando  $x$  receba  $a$  antes de receber  $b$ , já que, para o assunto da busca, aquele documento deve ser (supostamente) mais relevante que este. É esta a idéia que preside o ordenamento do resultado de uma busca nos *search engines*. O que os projetistas desses dispositivos fazem é elaborar uma hipótese interpretativa sobre o sentido (expressão aqui empregada no “sentido” bakhtiniano) da pergunta elaborada pelo usuário no diálogo de busca. É com base nessa hipótese interpretativa que os dispositivos avaliam as entradas encontradas nos seus índices e ordenam o resultado das buscas.

O critério clássico para classificação e ordenação dos resultados das buscas parte da premissa de que a incidência da expressão de busca num documento é indício do interesse do usuário. Assim, a taxa de incidência num documento e a exatidão da expressão são os critérios principais utilizados no ordenamento. Quanto mais ocorrências da expressão buscada (e quanto maior sua exatidão) forem encontradas num documento, maior será a probabilidade de ele ser ordenado nas primeiras posições apresentadas na saída. Esse critério elementar é refinado continuamente pelos dispositivos dedicados à busca, incluindo elementos “contextuais” e heurísticas complexas para a tomada de decisão classificatória; no entanto, a idéia geral é (quase) sempre esta. Duas são as hipóteses (relacionadas) utilizadas para dar conta da “intenção” do usuário: (i) Quando interroga um bibliotecário (o equivalente pré-ciberespaço do dispositivo de busca) a propósito de um livro, ele geralmente se refere ao “assunto” principal abordado no texto e (ii) existe boa probabilidade de que o tema de um documento se exprima através da incidência do mesmo vocabulário no corpo do texto.

Um critério simples de classificar a relevância elaborado com base nessa hipótese interpretativa poderia ser, por exemplo, determinar a razão entre a expressão tema e a quantidade total de palavras no documento. No nosso caso (até o parágrafo anterior contamos 1.500 palavras), a razão da relevância de alguns temas importantes neste documento – dispositivos de busca (ou *search engines*), ciberespaço e Internet – seria, respectivamente, 0,00467, 0,00400 e 0,00467, o que parece consistente com a estrutura temática deste

fragmento de documento. Poderíamos, por exemplo, supor que existe boa probabilidade de o tema do documento ser expresso na primeira linha (o título) e ponderar o cálculo de acordo; se arbitrarmos peso três para essa condição, o valor calculado de relevância naquela seção do documento da expressão “dispositivos de busca” (no original em inglês) saltaria para 0,01401. Esse critério pode ser refinado continuamente através de novas hipóteses sobre como os produtores de texto exprimem do ponto de vista vocabular o assunto dos seus documentos, agregando características típicas da Internet. Por exemplo, é altamente provável que um *site* cujo nome de domínio seja <http://www.searchengine-watch.com> se refira a dispositivos de busca! Isso pode se refletir como valor mais alto na ponderação das nossas entradas de índice.

Esse critério tem um problema: ele não reage como um bibliotecário humano culto e bem informado. Suponhamos que estamos utilizando um dispositivo de busca construído com base no critério descrito no parágrafo anterior para o cálculo da relevância dos documentos. Para uma busca pela expressão “ciberespaço” ele devolveria o presente texto antes do artigo A emergência do ciberespaço e as mutações culturais (ver <http://empresa.portoweb.com.br/pierrelevy/aemergen.html>), de Pierre Lévy, cujo cálculo de relevância para essa expressão de busca teria sido de 0.00110. É de se presumir que um bibliotecário humano, culto e bem informado, não respondesse desse modo, mas devolvesse o texto de Lévy em primeiro lugar. Afinal, este bibliotecário hipotético provavelmente imaginaria que estamos em busca de idéias de pesquisadores importantes e seu conhecimento de mundo o forçaria a se perguntar: mas afinal, quem é mais importante, um obscuro pesquisador júnior de uma universidade brasileira ou o filósofo francês Pierre Lévy?

Foi raciocinando deste modo que alguns pesquisadores de Stanford (Larry Page, Serguei Brin, Rajeev Motwani e Terry Winograd) elaboraram há alguns anos um critério distinto, batizado como *PageRank*. A idéia central daqueles pesquisadores é buscar um meio de determinar a “importância” de um documento na *web* consistente com o raciocínio daquele bibliotecário hipotético. Para isso, eles partiram da premissa de que um dos critérios acadêmicos de avaliação dos pesquisadores – a contagem de citações recebidas – seria aplicável para o cálculo da “importância” de um documento na *web*, tirando proveito da própria estrutura do hipertexto, baseada em *links* entre páginas. Para eles, “páginas [*web*] para as quais muitos *links* apontam são, em geral, mais ‘importantes’ que páginas com poucos *links*”, apontando para elas – do mesmo modo que “a simples contagem de citações tem sido utilizada para especular sobre os futuros ganhadores do Prêmio Nobel” (PAGE, 1998: 3). Pelo critério, a importância de uma página se propaga do mesmo modo que uma citação elogiosa a outro feita por um pesquisador muito importante. À época, esses pesquisadores determinaram,

por exemplo, que a página principal do Yahoo era (então) o documento para o qual mais *links* apontavam. Pelo critério adotado, eles admitem então que “se uma página tem um *link* saindo da *home page* do Yahoo, ele pode ser apenas um *link*, mas trata-se de um muito importante”, já que “esta página deveria receber um *ranking* mais alto que muitas páginas com mais *links* [apontando para elas], mas oriundos de lugares mais obscuros” (PAGE, 1998: 3).

Embora não seja dito no artigo citado, parte-se da premissa de que o usuário está interessado primariamente nos *sites* mais populares que atendam ao critério de busca. Neste caso, as saídas passam a ser classificadas e ordenadas segundo a taxa de referência ao documento feita por outros documentos. Quanto mais *links* forem encontrados em *sites* diferentes apontando para um determinado documento, maior a probabilidade de este aparecer nas primeiras saídas do resultado da busca. Trata-se aqui de determinar a audiência de um *site* a partir das referências feitas a eles por outros *sites* e devolver ao usuário aqueles documentos de maior audiência que atendam à expressão de busca utilizada.

É preciso afirmar inequivocamente que tudo indica que o critério de relevância do Google é o melhor disponível, não tanto pelas suas qualidades intrínsecas, mas pela palavra final dos usuários das buscas: a cada duas buscas realizadas no mundo, uma é feita neste dispositivo, a outra ficando por conta de pelo menos treze dispositivos, para contarmos somente os mais populares. Mesmo sem conhecer a natureza dessa diferença, os usuários devem ter escolhido o Google por força da consistência do produto das buscas realizadas no dispositivo com seus próprios interesses. Note-se que não é possível atribuir essa predominância simplesmente a um maciço investimento de capital: o Google nasceu pequeno, como “mera” pesquisa acadêmica – e só recebeu investimentos de capital depois de se tornar um sucesso de público. Apesar dessa maciça adesão, precisamos considerar as características intrínsecas do seu critério de relevância e indagar o que resulta dele: quais suas implicações para a prática social em torno de um dos eventos comunicativos mais comuns no ciberespaço. A nosso ver, as premissas assumidas por Page e colegas implicam em alguns problemas importantes.

Em primeiro lugar (e esse ponto é essencial para o restante do raciocínio), é preciso considerar que o cálculo da relevância relativa dos vários documentos que atendam a uma expressão de busca, ainda que imaginado somente como critério de ordenação dos resultados, é, na prática, um critério de seleção de que subconjunto do conjunto de documentos obtidos numa busca será, de fato, exibido ao usuário. Já dissemos que o produto da maioria das buscas, por envolver geralmente milhares de documentos, é impossível de ser processado pelos seres humanos. E os projetistas dos dispositivos sabem disso! Por exemplo, a busca pela

expressão "theory of games" no Google resultou, em 29 de outubro de 2006, em 404 mil documentos; no entanto, apenas 710 foram exibidos pelo dispositivo. A mesma expressão no Yahoo, resultou em 120 mil documentos encontrados, mas apenas mil exibidos! Isso é consistente com o comportamento de fato dos usuários. Como já assinalamos, apenas 12% dos usuários insistem na busca além da terceira página de resultados (IPROSPECT, 2006: 5). E mesmo que essa minoria de usuários pacientes insista na sua pesquisa, os dispositivos não lhe permitirão seguir além de um certo ponto arbitrário. Portanto, o cálculo da relevância relativa (não importando segundo qual hipótese foi elaborado) é utilizado de fato para a seleção do que será exibido ao usuário.

Ora, como vimos, a hipótese do Google é baseada no critério acadêmico de contagem de citações recebidas para julgamento da importância de um pesquisador. Admitamos, por hipótese, a validade desse critério no domínio para o qual elaborado. Trata-se de um julgamento pelos próprios pares do pesquisador. Embora esse julgamento nem sempre possa ser “justo”, ao longo do tempo as melhores pesquisas acabam por ser reconhecidas pelo menos por parte da comunidade acadêmica. Pensemos no caso de Galileu, quase condenado à fogueira por seus pares, um destino que – pelo menos metaforicamente – ameaça todas as idéias revolucionárias. Se hoje fizer minhas as críticas dos seus detratores, provavelmente serei eu o condenado à fogueira. Ocorre que no domínio acadêmico, a contagem de citações é apenas um critério de “ordenamento”, podendo servir para determinar o próximo ganhador do Prêmio Nobel, mas não sendo utilizada para decidir que pesquisas serão ou não submetidas à comunidade acadêmica. Esta dispõe de inúmeros mecanismos para assegurar a circulação de novas idéias, mesmo que produzidas por pesquisadores obscuros, oriundos de instituições igualmente obscuras, em qualquer parte do mundo. Além disso, essa comunidade é suficientemente pequena para que tais mecanismos sejam viáveis: sempre posso saber com pouco esforço o que outros pesquisadores em qualquer parte do mundo no meu domínio de interesses estão publicando, já que são relativamente poucas e facilmente acessíveis as fontes de informação a consultar.

Isso não ocorre na Internet, quando a consideramos isoladamente. Aquilo que a torna avessa aos controles e à censura – o fato de ser uma rede sem centro virtualmente ilimitada – é também aquilo que restringe o acesso à informação publicada: tudo pode estar *disponível*, mas pouco está efetivamente *acessível*. A circulação de idéias em grande escala na Internet depende consideravelmente dos dispositivos de busca – e se estes selecionam mais que ordenam o que estará acessível aos usuários, funcionam na prática como mecanismos de bloqueio à circulação de algumas informações, restringindo o que os usuários podem

conhecer. Por si só (e independente de qualquer ação institucional), os dispositivos controlam, censuram a circulação da informação na Internet – e o que eles censuram está diretamente relacionado às hipóteses interpretativas adotadas para elaborar o critério de cálculo da relevância relativa dos documentos buscados.

Como vimos, o critério do Google é baseado na audiência dos *sites*: ele valoriza as páginas na proporção direta da quantidade de *links* que apontam para ela, propagando esse valor para as páginas “citadas”. Isto significa que uma busca no Google retorna somente os *sites* mais populares para uma dada expressão de busca. Ora, idéias novas e revolucionárias não podem ser populares simplesmente porque ainda são novas e revolucionárias... Essas são as idéias censuradas pelo Google! Acrescente-se a isso aquelas idéias não tão novas e reacionárias, mas impopulares (admitamos que elas existam), sobre um determinado assunto e concluiremos que o Google é avesso aos direitos de expressão das minorias. A “missão” do Google – “organizar a informação mundial e torná-la universalmente acessível e útil” (GOOGLE, 2006) – precisa ser vista sob esta ótica menos otimista.

É fácil imaginar um experimento visando determinar se estamos ou não sofrendo de uma Síndrome de Cassandra ideológica. Suponhamos que estamos na metade do Século XVI e que um certo Nicolau Copérnico, obscuro astrônomo polonês, acaba de publicar um curioso estudo intitulado *De revolutionibus orbium coelestium* no hipotético *website* <http://www.frombork.edu> na não menos hipotética Internet da Renascença. Imaginemos também que isso é tudo o que, naquele momento, os astrônomos têm à disposição para divulgar suas idéias. Como se trata de um novo estudo de um pesquisador obscuro de uma universidade obscura com idéias excessivamente diferentes daquelas admitidas não apenas pelos seus pares, mas por todos os usuários da Internet Renascentista, nenhum outro *website* se deu ao trabalho de indicá-lo com um *link* de qualquer tipo. Estamos supondo também, para simplificar o problema, que o *website* <http://www.malleusmaleficarum.org>, àquele tempo, não publicasse listas exaustivas de candidatos à fogueira, mas apenas daqueles efetivamente incinerados para edificação dos fiéis. Como não existe nenhum *link* apontando para ele, o dispositivo de busca <http://www.googlorum.com> classifica o documento com a menor relevância possível. O único modo de alguém interessado nos movimentos celestes localizar esse documento no Googlorum é digitar uma combinação de palavras que só exista naquele texto e em nenhum outro com relevância maior, o que não é plausível, a menos que o usuário tenha dons de clarividência, caso em que não teria necessidade de utilizar o dispositivo de busca. Num conjunto universo de pelo menos 8 bilhões de documentos (é esta a quantidade de documentos que o Google alega ter já indexado) e milhões de *websites* (somente o Brasil



dispõe de pelo menos um milhão de domínios registrados), a probabilidade de um usuário, navegando ao acaso, localizar um documento específico sem a utilização de um dispositivo de busca é, para todos os efeitos práticos, desprezível. Nesse cenário, o documento de Copérnico está irremediavelmente inacessível – e nós condenados a continuar acreditando ser a Terra o centro do universo, circundada por uma esfera de estrelas fixas.

Esta é, certamente, uma situação limite, até porque a Internet fornece um conjunto de outros mecanismos de acesso à informação que permitem a circulação de idéias novas, embora não na mesma escala e com a mesma abrangência dos dispositivos de busca. No entanto, o que queremos assinalar é essa possibilidade – assustadora – implícita naquela ferramenta que muitos parecem considerar o oráculo dos novos tempos. Se assim for, a sacerdotisa só interpreta uma única profecia – a da uniformidade. Infelizmente, nada temos a contrapor a esta concepção para o cálculo da relevância, em particular quando todos parecem satisfeitos com ela.

#### Referências Bibliográficas

- BAKHTIN, M. *Estética da criação verbal*. Tradução do russo por Paulo Bezerra. 4. ed. São Paulo: Martins Fontes, 2003.
- FOUCAULT, M. *A ordem do discurso*. Aula inaugural no Collège de France, pronunciada em 2 de dezembro de 1970. Tradução do francês de Laura Fraga de Almeida Sampaio. 10.ed. São Paulo: Loyola, 2004.
- FAIRCLOUGH, N. *Discurso e mudança social*. Tradução do inglês de Izabel Magalhães et al. Brasília: Editora Universidade de Brasília, 2001.
- GOOGLE corporate information: company overview. *Google*. Disponível em: <<http://www.google.com/intl/en/corporate/index.html>>. Acesso em: 29 outubro 2006.
- IPROSPECT. *Search engine user behavior study*. 2006. Disponível em: <[http://www.iprospect.com/premiumPDFs/WhitePaper\\_2006\\_SearchEngineUserBehavior.pdf](http://www.iprospect.com/premiumPDFs/WhitePaper_2006_SearchEngineUserBehavior.pdf)>. Acesso em: 27 Agosto 2006.
- PAGE, L; BRIN, S; MOTWANI, R; WINOGRAD, T.. *The PageRank citation ranking: bringing order to the web*. Stanford University, 1998. Disponível em: <<http://dbpubs.stanford.edu:8090/pub/1999-66>>. Acesso em: 5 outubro 2005.
- SULLIVAN, D. Search Engine Size Wars V erupts. *Search Engine Watch*, 11 Novembro 2004, Search Ratings & Stats. Disponível em: <<http://blog.searchenginewatch.com/blog/041111-084221>>. Acesso em: 28 Outubro 2006.
- SWALES, J.M. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press, 1990.