# PALAVRAS, PALAVRAS DESLOCADAS PARA UM SIGNIFICADO

## *WORDS, WORDS SHIFTED TO A MEANING*

**Aldo de Albuquerque BARRETO**
Pesquisador do Instituto Brasileiro de Informação em Ciência e Tecnologia
Pesquisador Sênior do Conselho Nacional de Desenvolvimento Científico e Tecnológico
*e-mail*: aldo.barreto@gmail.com

**Resumo**
O artigo revisa e estuda a estrutura do texto escrito e sua análise morfológica com a finalidade de, operando de maneira automatizada, extrair informações tanto para a elaboração como para a representação do significado de documentos. Fornece subsídios técnicos e teóricos para construção de uma configuração para análise das palavras numa narrativa, sua freqüência e sua relação com o significado. Representa, teoricamente, reflexão para apreciação da manifestação do fenômeno da informação, aqui entendido, como a sensibilidade na percepção do conteúdo semântico das estruturas de informação pelos sentidos e pela consciência.
**Palavras-chave**
Indicadores de significação. Freqüência de palavras. Análise computacional do português. Relevância por análise morfológica. Indicadores de relevância. Indicadores da nuance do texto.

***Abstract***
*This paper is oriented to the morphologic study of the structure of the written text and its analysis with the purpose of extracting words and counting its frequency for strategic uses of the information content of a document. In conceptual terms, it represents a reflection for appreciation of the manifestation of the phenomenon of the information, as the sensibility in the perception of the semantic content of the structures of information for the senses and for the conscience.*
***Key words***
*Text content indicators*. *Word frequency counting*. *Computer analysis of natural languages*. *Morphological text analysis*. *Relevance indicators. Appreciation of text nuance.*

## INTRODUÇÃO

A informação, quando referencia o homem lúcido[1] ao seu destino, participa de seu caminho vivencial ao estabelecer referências para percorrer sua odisséia individual no espaço e no tempo. Associada ao conceito de ordem e de redução de incerteza identifica-se com a organização dos sistemas de seres vivos racionais. Neste sentido, este artigo somente traz reflexão da qualidade do fenômeno, quando entre seres humanos, onde existe emissor, canal de transferência, código de registro comum e destinatário.

[1] A lucidez é um dom e um castigo. Está tudo numa palavra. Lúcido vem de Lúcifer, o arcanjo rebelde, o Demônio. Lúcifer é também o luzeiro do amanhecer, a primeira estrela, a que mais brilha e a última a se apagar. Lúcido vem de Lúcifer, Lúcifer, de Lux e Ferous, que quer dizer: aquele que tem luz ou gera luz ou permite a visão interior. Deus e Demônio, tudo junto. O prazer e a dor. Lucidez é dor, e o único prazer que se pode conhecer, o único que se parece remotamente à alegria é o prazer de permanecer consciente da própria lucidez. O silêncio da compreensão, o silêncio do simples estar. E nisto se vão os anos, nisto se foi à bela alegria animal. (PIZARNIK, Alejandra. Disponível em: <http://www.cibernetic.com/ALE/index.html>. Acesso em: 28 abr. 2010).

A essência do fenômeno da informação se efetiva entre emissor e receptor, quando acontecem transferência e apropriação de conhecimento. Adequadamente assimilada, a informação modifica o estoque mental de saber do indivíduo e traz benefícios para seu desenvolvimento e da sociedade em que vive. A questão que se coloca é trabalhar com a informação no que se refere à tipologia da estrutura de suporte, considerando sua ingerência na produção do conhecimento.

A produção da informação segue um processo de transformação com ações definidas e se apóia em procedimentos orientados por uma racionalidade específica. Como precursora da intenção de provocar conhecimento no indivíduo e na realidade, a informação pode ter diferentes alicerces de registro e trilhar variados fluxos relativos à sua administração e à sua distribuição.

A estrutura de informação é concebida como qualquer base[2] de inscrição de informação que o aceite como tal, ou seja, um conjunto de elementos que formam um todo ordenado na narrativa com seguimento e finalização coesa de enunciados. Os alicerces de informação mantêm variada tipologia. Um texto é um conjunto de expressões inscrito numa base, com multiplicidade de configurações de uma língua. Constitui um todo unificado passível de ser distribuído por um canal de transferência. Seu discurso de significação é uma elaboração do autor. Porém, quando distribuído, o texto associa, em sua amplitude, a leitura, o receptor e sua interpretação ou reconstrução. É feito de escritas múltiplas e de várias culturas que entram em diálogo e contestação e que se acumulam no leitor. Neste, está o ambiente exato em que se inscrevem todas as referências, das quais, uma escrita é feita. A unidade do texto não está em sua origem, mas em seu destino. Tal destino não é pessoal: o leitor é um homem sem história, sem biografia e sem psicologia.

Uma estrutura de informação pode ser linear, seqüencial e centrada em narração contínua. Um texto pode, também, ser acêntrico e sem destino certo, composto de várias estruturas que se narram em paralelo. A escrita, ao tempo em que concede ao homem valores visuais, ocasiona uma consciência fragmentada, ao contrário da convivência nos espaços auditivos, onde o convívio de enunciados multivariados, com muitas vozes simultâneas, é mediado pelo espaço e pela distância. Sob este prisma, a tipografia exterminou a cultura auditiva tribal. Permitiu à cultura escrita multiplicar possibilidades de se enunciar no tempo e no espaço. O homem, com seu pensamento linear e seqüencial, qualifica, organiza, classifica e assimila suas informações de modo hierárquico, numa série contínua de graus e escalas, que comporta famílias temáticas, em ordem crescente ou decrescente; organização por classes, com indicação de subordinações em relação à herança do universo particularizado das palavras.

Assim, a estrutura de informação constitui evento privado em sua criação e se complementa em tempo finito. Sua circulação e transferência, por sua vez, ocorrem no espaço público, para um número indefinido de leitores. Mas, todo ato de interpretação e apropriação é

[2] Base: local da inserção das inscrições de informação, que definem o modo da estrutura a que pertence. Um dos desafios mais intrigantes na biofísica é a questão da memória. Sabemos que temos vários bilhões de neurônios, além de outras células localizadas no cérebro (as células gliais, por exemplo). Os neurônios, por sua vez, estão espalhados pelo corpo (com maior concentração no cérebro) e são responsáveis, em parte, pela transmissão de estímulos sensoriais. Um neurônio é constituído por um corpo celular, dendritos ("ramificações" que partem desse corpo) e uma "cauda" (extensão), o axônio. Quando estimulado, ele produz uma diferença de potencial que gera uma tênue corrente elétrica. Esse estímulo elétrico se propaga e permite que libere substâncias específicas (neurotransmissores) que fazem o contato dele com neurônios vizinhos, formando sinapses. O conjunto de células interligadas forma uma rede que mantém semelhança com redes de sistemas físicos. Não se sabe como, mas a rede de neurônios tem a capacidade de gerar informação. Temos lembranças graças a esse complexo sistema de células. O estudo do cérebro e de suas redes (redes neurais) tem contribuído para o desenvolvimento de sistemas de informação que podem levar a um computador biológico. Fonte: Folder do CBPF / MCT.

uma condição privada e de solidão fundamental, em que o pensamento se refugia no âmago de cada privacidade.

A apropriação da informação revela um ritual de interação entre sujeito e determinada estrutura de informação, que gera (no sujeito) modificação de suas condições de entendimento e de saber acumulado. Tal apropriação representa um conjunto de atos voluntários, pelo qual o indivíduo reelabora seu mundo, modificando seu universo de conteúdos simbólicos. É uma criação em convivência com suas cognições prévias e com sua percepção. É início de algo que nunca iniciou antes, mas que resulta, sempre, numa modificação como conseqüência do ato em si, embora possa ocorrer um retorno para permanência ao seu estado inicial de saber.

A assimilação da informação consiste em condição imprescindível ao receptor para validar a informação acessada. Não é suficiente que a mensagem seja intencionalmente planejada na distribuição e para o acesso. O conteúdo deve atingir espaços semânticos compatíveis e harmoniosos para sua compreensão e aceitação.

Consideram-se dois estados de consciência relativos ao processo que media a informação para conhecimento: (**1**) através de um pensamento convergente; (**2**) através de um pensamento divergente (GUILFORD, 1959), com o adendo de que se atribui esta opção para as diferentes estruturas com que o ser humano lida, no cotidiano. Quando se percebe que o texto alcança o conhecimento via pensamento convergente não se exclui a possibilidade de pensamento divergente no processo. A intenção é frisar o tipo focal de pensamento que uma ou outra estrutura induz na formação do conhecimento.

Aliás, pensamento convergente é aquele em que o enunciado na estrutura se direciona a uma cadeia de ligações cognitivas precisa e direcionada a um ponto. É o pensamento determinado, convencional e pontual, que se abriga no interior de uma composição. Pensamento divergente, por seu turno, é aquele em que a estrutura de informação induz a um caminhar cognitivo em diferentes direções, como que pesquisando livremente os meandros dos documentos entrelaçados, com múltiplas escolhas de novos caminhos antes de desenvolver uma ligação com apropriação final. Logo, pensamento divergente é aquele em que a seleção das palavras para o texto caminha em diferentes direções como que pesquisando livremente os meandros das figuras de elaboração do estilo no momento da edição da informação.

Para Foucault (1992, paginação irregular), o que denomina como "função-autor" não se constrói simplesmente atribuindo um texto como sendo de um indivíduo, com poder criador, mas constitui “característica do modo de existência, de circulação e de funcionamento de alguns discursos no interior de uma sociedade". Ou seja, indica que tal ou qual discurso deve ser recebido de certa maneira e que deve, em determinada cultura, receber certo estatuto. O que faz de um indivíduo um autor é o fato de, através de seu nome, ser possível delimitar, recortar e caracterizar os textos que lhes são atribuídos.

A narrativa é exposição de um acontecimento ou parte de um acontecimento encadeado, real ou imaginário por meio de palavras, imagens ou texto de narração. Afirma-se que a geografia das narrativas é onde os documentos mantêm um limite territorial, isto é, quando há um ponto de referência exato de jurisdição. A evocação simbólica é operada por associações e referências do passado e projeções do futuro. É limitada, unicamente, pela riqueza das estruturas de memória ativadas. O significado do texto está conectado à relação entre informação e estado da memória do receptor, seu conteúdo e seus contextos. Na interpretação da informação, o receptor fica liberado da intenção do emissor.

**AS PALAVRAS E O SIGNIFICADO DO TEXTO**

A palavra é a menor unidade de uma estrutura significante, com condições representacionais e se agrega, formando frases, sentenças e sistemas de sentenças. Totaliza-se no texto, considerado como um conjunto simbolicamente significante. Em outras palavras, texto é um conjunto de palavras e frases articuladas, escritas sobre qualquer suporte formado por enunciados e narrativas que compõem uma peça de informação sobre certo tema. Trata-se de um conjunto coerente, em que, morfologicamente, as palavras podem falar do texto. Estruturalmente, o estudo do arcabouço e o conjunto de palavras numa narrativa constituem uma questão da morfologia do texto e permite decompor esta estrutura em suas unidades mais simples de significado.

A história do estudo de palavras numa narrativa se mistura com a própria história da Ciência da Informação (CI), quando, em 1945, com o fim da Segunda Guerra Mundial, a informação mantida secreta naquele período é colocada à disposição do mundo. Vannevar Bush (ver http://pt.wikipedia.org/wiki/Vannevar_Bush), responsável pelo Comitê Nacional de Pesquisa do *Office for Scientific Research and Development*, criado pelo presidente Roosevelt nos Estados Unidos da América, aproveitando sua experiência no cargo e sua convivência com as informações geradas como subproduto do trabalho dos cientistas, no ano citado, escreve *As we may think* (ver http://www. theatlantic.com/doc/194507/bush). Consiste em obra que discute a explosão da informação em ciência e tecnologia (C&T) advinda do período pós-guerra e dos possíveis entraves para organizar e repassar à sociedade o conhecimento gerado.

O artigo de Bush aparece pela primeira vez, em 1939, numa carta ao editor da revista *Fortune* (ver http://people.ischool.berkeley.edu/~buckland/goldbush.html), mas sua versão histórica está em *Atlantic Monthly* e, posteriormente, a renomada revista *Life* publica várias observações e chamadas acerca do trabalho. Nessa época, se reforça a noção de associação de conceitos ou palavras na análise e na organização dos textos e, conseqüentemente, da informação. Associar palavras é vista como o padrão que o cérebro humano utiliza para transformar conteúdos em conhecimentos.

Assim, os processos para armazenar e recuperar informação são considerados mais eficientes se operacionalizados por associação de conceitos – *As we may think*: como nós pensamos. Tais idéias são muito novas para a época e, portanto, provocam tal *frisson* que refletem na intelectualidade londrina. Em 1946, um ano após o término da Segunda Guerra Mundial, a *Royal Empire Society Scientific* realiza Conferência para discutir os impactos do fluxo da informação. Em 1948, se dá a Royal *Society Scientific Information Conference* (ver http://www.ncbi.nlm.nih.gov/pmc/articles/PMC194801/? page=1), também em Londres, Inglaterra. Naquele momento, cerca de 340 cientistas de todo o mundo se reúnem para reportar os problemas provenientes do acúmulo de informações em sua área de conhecimento. Nasce, assim, outra área voltada para o gerenciamento de conteúdos em C&T.

Um ano após a Conferência da *Royal Society* de Londres, cientistas se reúnem com o intuito de instalar o *Institute for Information Scientists* (IIS, ver http://www.iis.org.uk) com a pretensão de se institucionalizarem como cientistas. Nessa mesma época, em 1952, um grupo de cientistas da informação cria o *Classification Research Group* (ver http://web.archive.org/web/20011222083409/alexia.lis.uiuc.edu/ review/summer1995/spiteri.html) para propor teorias inovadoras para armazenar e recuperar a informação, haja vista que o problema de então é o grande volume de informação e sua gestão. Os profissionais que fundam o IIS, instalam, ainda, o primeiro curso de pós-graduação em Ciência da Informação em *The City University*, Londres.

Os cientistas que se envolvem com a nova forma de trabalhar a informação, temerosos de perder seu *status* acadêmico, denominam a nova área de CI, a qual, na verdade, ainda não comporta, à época, um conjunto de conhecimentos socialmente adquiridos ou produzidos, historicamente. Nem é dotada da universalidade e da objetividade inerentes a um arcabouço teórico consensual que permita sua transmissão nem possui um conjunto estruturado com métodos, teorias e linguagens próprias, visando compreender e orientar sua natureza, com a ressalva de que essas observações se referem à CI e não à secular Biblioteconomia. Nasce, então, uma atividade social aplicada, com a denominação de ciência.

Nesse início, tenta-se imprimir um caráter científico à CI mediante as supostas Leis da Bibliometria, que, de fato, se configuram como:

> [...] a aplicação de técnicas estatísticas e matemáticas para descrever aspectos da literatura e de outros meios de comunicação (análise quantitativa da informação). A bibliometria foi originalmente conhecida como "bibliografia estatística" (termo cunhado por Hulme, em 1923), sendo o termo "bibliometria" criado por Otlet, em 1934, em seu *Traité de Documentation*. Contudo, o termo apenas se popularizou em 1969, a partir de um artigo de Pritchard que discutia a polêmica bibliografia estatística ou bibliometria?. (ARAÚJO, 2006, paginação irregular).

As medições estatísticas mais conhecidas são o método de medição da produtividade de cientistas de Lotka (1926), a lei de dispersão do conhecimento científico de Bradford (1934) e o modelo de distribuição e freqüência de palavras num texto de Zipf (1949).

As Leis de Zipf (GUEDES; BORSCHIVER, 2005) estão relacionadas com a freqüência de ocorrência de palavras num texto e diretamente pautada com seu conteúdo de informação. Para Zipf, num texto suficientemente longo, apenas um pequeno número de palavras possui freqüência muito alta, inevitavelmente relacionada com o tema em questão. É interessante notar a proximidade no tempo dos trabalhos de Vannevar Bush (1945) e George Kingsley Zipf (1949), ambos tratando de palavras para representar os conteúdos de um documento.

A estrutura do texto escrito e sua análise morfológica são construídos e desconstruídos pela palavra. *A* "estrutura" é pensada como elementos que formam um todo ordenado e com princípios lógicos, com coerência de raciocínio e de idéias. Um texto de informação possui características estruturais de linguagem que admitem análise morfológica e onde partes podem representar o todo.

Uma narrativa, quando construída, vem de uma linguagem simplificada do pensamento de seu gerador para uma linguagem determinável de edição da escritura. Walter J. Ong (1988, paginação irregular) discute a maneira como a escrita distancia o autor de seu pensar. Indica as características da linguagem do pensamento de onde o autor organiza mentalmente sua narrativa, antes da escrita:

> [...] as expressões são aditivas em sua narrativa, não se subordinam; possui uma tendência para ser redundante ou a reutilizar conceitos constantemente; uma organização simples em sua forma, com frases pequenas e com palavras curtas; possui uma tendência à estabilidade interna com um retorno constante aos conceitos já usados; é situacional mais que abstrata.

Por outro lado, estudos realizados por Lancashire (1993) permitem afirmar que a escrita de edição adota códigos delineáveis. É explícita, formal, de padrões normativos e de procedimentos formalizados; procura eliminar as repetições de conceitos, a redundância e as palavras indeterminadas; utiliza figuras de linguagem e é rica em metáforas; usa uma grande

fluência de termos e extrema liberdade semântica; utiliza estruturas sintáticas complexas, mas determináveis e possivelmente generalizáveis. Ainda podemos dizer que lança mão de excesso de sinonímia e de conectores entre conceitos; utiliza as palavras sem preocupação com seu tamanho, em frases de construção livre; e, também, é uma linguagem morfologicamente coerente e passível de manter alguma definição de padrões e procedimentos.

Assim, é possível identificar, a partir da linguagem de edição, a escrita usada pelo autor no texto, e dentro de um contexto de análise, afirma-se que:

- A qualidade de relevância do conteúdo de um texto está diretamente relacionado com a quantidade das palavras mais freqüentes numa contagem de todas as palavras daquele texto.

- A nuance de uma narrativa está diretamente conectado com as palavras de freqüência igual a um que aparecem como holofrases (palavras que implicam o significado único e inteiro) nesta narrativa. São as palavras que aparecem somente uma vez e, quando devidamente elaboradas, mostram a singularidade de um texto em estudo. Denotam a nuance em forma de etiquetas. Como um pensamento a elaborar seu discurso inicial quando um autor como que busca palavras específicas para formar seu enunciado, mas quer colocar uma nuance que o diferencie de outros textos. As coisas escritas são, agora, cada vez mais tecnicizadas e o particular só se encontra, hoje, nas nuances que um autor deixa como pista.

- A informação, apontada por sua nuance de singularidade, pode indicar uma condição de prioridade para determinado receptor. Assim, as palavras que indicam a relevância de um texto, as de maior freqüência, situam-se num conjunto que revela seu conteúdo.

- As palavras de freqüência igual a um se agregam em conjuntos de significatividade, em função do tamanho das palavras. A palavra, aqui, tem conexão com uma holofrase ou um unitermo, ao denotar uma nuance. Quanto maior a palavra única, maior sua densidade para estar mostrando uma nuance deixada pelo texto gerador.

- A análise das palavras de maior freqüência de um texto operando para representação de seu conteúdo num procedimento para exame e indexação não consiste em novidade. O tema foi examinado por George Kingsley Zipf (1902-1950), professor de Lingüística em Harvard e publicado, em 1932, em seu *Selective studies and the principle of relative frequency in language,* no que se convencionou chamar de a Lei de Zipf.

Hoje, porém, de acordo com Santos (2010), este tipo de exame se fundamenta num contexto teórico e metodológico chamado análise computacional do português como linguagem natural.

## AS PALAVRAS QUE DENOTAM A NUANCE DA NARRATIVA

A análise das palavras que aparecem com freqüência no texto mostra sua singularidade ou individualidade, o que constitui condição nova para a qualidade na análise automatizada do

conteúdo. Convenciona-se chamar tal condição de nuance ou prioridade graças à singularidade da relação autor **x** texto e que, junto com as palavras de maior freqüência, fornecem representação morfológica do texto em relação à sua significância, relevância e prioridade.

As palavras numa narrativa que aparecem apenas uma vez indicam um pensamento livremente divergente. As palavras que aparecem com freqüência maior do que um é o resultado do viés cognitivo de quem escreve no formalismo pontual de um pensamento convergente. Esta configuração de freqüência de palavras permite inferir o comportamento morfológico em diferentes áreas do conhecimento.

O **Quadro 1** mostra o comportamento para textos de informação, considerando a área do conhecimento e seu relacionamento entre as palavras que os compõem, com o adendo de que, no estudo, adota-se o artigo espaço virtual - mundo – real (ver http://docs.google.com/Doc?docid=0Af2oVw3ez502ZGR2NmQyODVfOTZjeHR0Zm1mYw&hl=en).

| ÁREA | Características do subcódigo lingüístico | Léxico de edição do texto | Cadeia de pensamento na edição do texto |
|---|---|---|---|
| **LEVES**<br><br>Ciências Humanas e Sociais | Difícil delimitação do subcódigo.<br><br>Extrema liberdade semântica.<br><br>Discurso longo, difuso e informal com conceitos interligados. | Estruturas sintáticas complexas.<br><br>Excesso de sinonímia, metáforas, conectores e plurais.<br><br>Linguagem sem controle ou inibição instrumental. | Fluência de idéias e palavras com independência em elaborar significados.<br><br>Revogação de palavras da memória com grande liberdade conceitual.<br><br>Pensamento divergente na recognição dos conceitos. |
| **INTERMEDIÁRIAS**<br><br>Ciências Sociais Aplicadas com *slant* para a tecnologia | Código mais formalizado e manifesto.<br><br>Condições semânticas controladas.<br><br>Discurso semi-técnico. | Estruturas sintáticas elaboradas.<br><br>Vocabulário com alguma inibição pela formalização de figuras de linguagem. | Alguma independência na elaboração dos significados.<br><br>Revocação de conceitos da memória com orientação ao assunto.<br><br>Predominância de pensamento de recognição convergente dos conceitos na memória. |
| **DURAS**<br>Ciências Exatas e da Natureza | Código formalizado e delineado.<br><br>Discurso técnico e direcionado. | Estruturas sintáticas simples. Vocabulário formal e específico.<br><br>Bastante inibição em relação às figuras de linguagem. | Pouca fluência de idéias e conceitos novos.<br><br>Precisão na revocação dos conceitos da memória.<br><br>Pensamento altamente convergente na revocação das palavras. |

**Quadro 1 –** Comportamento para textos de informação: área e relacionamento entre as palavras

## AVALIAÇÃO INDEPENDENTEMENTE DOS RESULTADOS ENCONTRADOS

No programa de Ciência da Informação da Universidade Federal do Rio de Janeiro, ano 2001, o pesquisador e autor deste artigo ofertou a disciplina *Estrutura e fluxos de informação* para alunos de doutorado e mestrado. Durante um semestre, discutiram as bases teóricas do

presente relacionamento de palavras num contexto de escrita e submeteram à prova o modelo de trabalho ora exposto. De fato, produziram excelente material complementar, aqui, aproveitado parcialmente.

Considerando os resultados do tratamento dos textos escritos na língua portuguesa, verificou-se que as freqüências distribuíam-se de forma a compor núcleos com determinadas características quanto ao grupo de suas palavras significativas. Identificou-se, no conjunto de palavras de maior freqüência, as que poderiam representar o conteúdo da narrativa, com uma freqüência média de repetição de oito a 10 vezes por uma mesma palavra. Neste mesmo campo, estão as palavras que, marcadamente, indicam o conteúdo com freqüência que poderia chegar a 150 repetições, a exemplo da palavra informação.

## A SINGULARIDADE DO CONTEÚDO

Com a ressalva de que ao longo do estudo, adotou-se um *software* de análise morfológica do conteúdo de textos em português como linguagem natural, no que se refere à singularidade do conteúdo, as expressões mais significativas (prioritárias para a compreensão do pensamento do autor) foram identificadas nas palavras de freqüência igual a um. Ali estava a nuance colocada pelo autor no texto.

As palavras de freqüência um aparecem em significativa quantidade nos textos numa proporção que varia de 50% a 60% do total de palavras da narrativa. Figuram em conjuntos de significância relacionados com o tamanho da palavra em número de letras. Estes *clusters* de palavras detinham de 15 a 19 palavras indicando a nuance do texto em análise. As palavras irrelevantes neste conjunto de freqüência um são eliminadas automaticamente pelo *software* adotado, atingindo precisão de 85. Em geral, são nomes próprios, adjetivos, advérbios, apostos, qualificativos, pronomes, preposições, conjunções, artigos, numerais (plurais), interjeições, estrangeirismos, data e local.

## A TENSÃO COGNITIVA DO USUÁRIO

Tensão cognitiva é o estresse provocado pelo exame de grande quantidade de informação e o tempo necessário para sua avaliação e potencial interiorização. Quando a informação é apresentada de maneira visualmente destinada à percepção, numa estrutura gráfica que permite visualização amigável, o esforço cognitivo é diminuído para o receptor, no processo de julgamento de valor e decodificação. A tranqüilidade cognitiva permite ao receptor lidar com nuance da informação, porquanto a sensação da percepção pode ser transmitida visualmente.

Além disto, levando em conta as observações anteriores alusivas ao trabalho com as palavras que formam a estrutura do texto, observa-se ser possível estabelecer zonas de informação, configuradas como prioritárias e relevantes, o que reduz a tensão cognitiva do receptor ao interagir com a informação. Com o zoneamento e o acoplamento de mecanismos que possam agregar informação, há decréscimo da tensão ora discutida. O usuário passa a ter acesso às informações não apenas relevantes e prioritárias às suas demandas, mas, também, àquelas que podem ser facilmente apropriadas por ele, segundo zona de qualidade intensa, agregando documentos que atendam às suas variadas necessidades informacionais.

O exemplo de tal pensamento pode ser apresentado assim: incorporam-se aos estoques de informação representações-indicadores, as quais agregam a informação em níveis de prioridade e de relevância de acordo com determinado perfil de necessidade. Exemplificando:

um produtor rural que busca solução para combater certa praga em sua lavoura acessará informações relevantes e prioritárias em relação ao seu perfil. Quer dizer, é essencial a apropriação adequada do conhecimento: um pesquisador, ao acessar o mesmo estoque em busca de informações sobre o mesmo problema, deve se defrontar com um universo de informações muito menor para decidir sobre a utilidade de cada item de informação.

Na verdade, é possível extrair diversas inferências a partir da desconstrução da narrativa em suas palavras que a compõem. O **Gráfico A** mostra a relação do número de palavras de maior freqüência num documento, reforçando que as palavras mais freqüentes estão num conjunto de freqüência cinco a 75, em média. O **Gráfico B**, por sua vez, sintetiza a relação das palavras de freqüência um, mostrando a importância do número de letras neste tipo de palavra, quanto maior a palavra maior sua significância.

GRÁFICO A

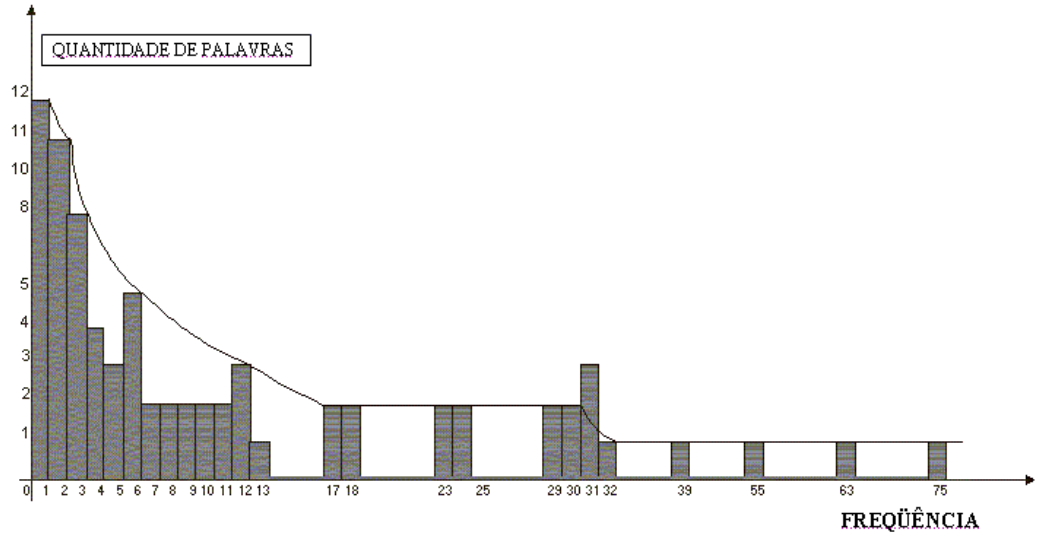
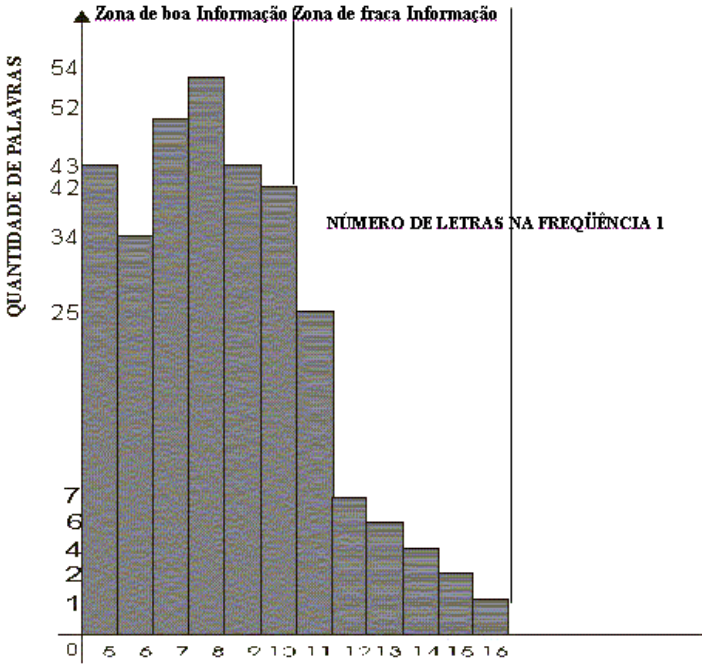



GRÁFICO B



Um exemplo de análise de um texto de informação por sua desconstrução em palavras para reconstrução de seu significado é o que se propõe fazer para ilustrar o discutido neste artigo. Primeiro, é preciso conhecer o texto *Mundo real – mundo virtual*, narrativa de sete laudas e disponível em <http://aldoibct.bighost.com.br/palavras.htm>. Por se tratar de apropriação de um texto produzido por uma autora portuguesa, muitas palavras aparecem com ortografia do português daquele País. Recorrendo-se, então, a um *software* de análise computacional do português como linguagem natural, se obtém um resumo das condições das palavras do conteúdo do texto (**Quadro 2**).

| DESCRIÇÃO | TOTAIS |
|---|---|
| Total de palavras | 3.498 |
| Total de verbos | 260 |
| Total de substantivos | 293 |
| Total de nomes próprios | 6 |
| Total de adjetivos | 113 |
| Total de advérbios | 49 |
| Total de pronomes | 76 |
| Total de preposições | 40 |
| Total de conjunções | 19 |
| Total de artigos | 23 |
| Total de numerais | 7 |
| Total de interjeições | 0 |
| Total de estrangeirismos | 4 |
| Total de outras palavras | 475 |

**Quadro 2 –** Características do espaço virtual – mundo real

A indicação preliminar é que se deve verificar se é possível reconstruir o significado do texto para assim obter análises subseqüentes com aproximadamente 293 substantivos, considerados "palavras fortes" determinando o sentido da narrativa. As 11 palavras mais freqüentes do texto constam do **Quadro 3**. Enquanto isto, as palavras que denotam a nuance do texto possuem freqüência igual a um no texto, como constam do **Quadro 4**.

| TAMANHO | LOCAL | FREQÜÊNCIA | PALAVRA |
|---|---|---|---|
| 10 | 201 | 17 | Comunidade |
| 8 | 173 | 16 | Internet |
| 10 | 356 | 14 | Informação |
| 9 | 329 | 11 | Realidade |
| 11 | 413 | 8 | Ciberespaço |
| 8 | 263 | 8 | Virtuais |
| 10 | 317 | 7 | Computador |
| 10 | 1.869 | 7 | Multimídias |
| 8 | 677 | 6 | Ambiente |
| 8 | 638 | 5 | Extensão |
| 10 | 36 | 5 | Tecnologia |

**Quadro 3 –** As 11 palavras mais freqüentes do texto

No **Quadro 4**, a primeira coluna refere-se ao número de letras da palavra. A segunda, ao posicionamento no texto e, por último, à freqüência com que a palavra aparece no texto. Ademais, é interessante conferir com o texto disponível, e os quadros indicam que o texto é sobre informação na e sobre a Internet, e, ainda, acerca de novas tecnologias, ambientes virtuais e novas tecnologias da informação. Seu viés é evidente pelas palavras que o autor coloca uma só no texto: inclusão digital, instrumentos e aplicações e sensações das tecnologias intensas de informação. Vale notar, ainda, que todas as palavras significantes possuem oito letras ou mais.

Para amostra do método ora introduzido, sintetizam-se as palavras mais freqüentes e que revelam a relevância do texto que o pesquisador escreve neste momento (**Quadro 5**).

| N LETRAS | POSIÇÃO | FREQÜÊNCIA | PALAVRA |
|---|---|---|---|
| 8 | 144 | 1 | Política |
| 8 | 985 | 1 | Terminal |
| 8 | 1.241 | 1 | *Clusters* |
| 8 | 1.296 | 1 | Estatuto |
| 8 | 1.367 | 1 | Romances |
| 8 | 1.649 | 1 | Motorola |
| 8 | 2.470 | 1 | Imprensa |
| 8 | 2.500 | 1 | Conteúdo |
| 8 | 2.793 | 1 | *Displays* |
| 8 | 2.867 | 1 | Sensação |
| 8 | 2.913 | 1 | Conflito |
| 8 | 3.124 | 1 | Infinito |
| 8 | 3.328 | 1 | Culturas |
| 8 | 3.458 | 1 | Educação |
| 9 | 436 | 1 | Biológico |
| 9 | 1.053 | 1 | Amputação |
| 9 | 1.624 | 1 | Celulares |
| 9 | 1.635 | 1 | Satélites |
| 9 | 2.095 | 1 | Documento |
| 9 | 2.221 | 1 | Navegação |
| 9 | 3.363 | 1 | Info-ricos |
| 10 | 353 | 1 | Digitaliza |
| 10 | 1.196 | 1 | Alucinação |
| 10 | 1.319 | 1 | Literatura |
| 10 | 1.986 | 1 | Totalidade |
| 10 | 2.445 | 1 | Hipermídia |
| 10 | 2.650 | 1 | Artificial |
| 10 | 2.725 | 1 | Território |
| 10 | 2.755 | 1 | Visualizam |
| 10 | 2.849 | 1 | Mobilidade |
| 10 | 3.365 | 1 | Infopobres |
| 11 | 909 | 1 | Componentes |
| 11 | 1.190 | 1 | Psicanálise |
| 11 | 1.215 | 1 | Matemáticos |
| 11 | 1.332 | 1 | Corporações |
| 11 | 3.283 | 1 | Consciência |
| 11 | 3.473 | 1 | Interativa |
| 12 | 1.005 | 1 | Telepresença |
| 12 | 1.146 | 1 | Cibercultura |
| 12 | 1.230 | 1 | Complexidade |
| 12 | 1.735 | 1 | Eletrônicos |
| 12 | 2.523 | 1 | Aumentativos |
| 12 | 2.727 | 1 | Cartografado |
| 14 | 664 | 1 | Tridimensional |
| 14 | 748 | 1 | Hiper-realidade |
| 14 | 1.085 | 1 | Transcendência |
| 14 | 3.460 | 1 | Entretenimento |
| 15 | 1.504 | 1 | Comunicacional |
| 17 | 1.861 | 1 | Multidimensionais |
| 18 | 900 | 1 | Representatividade |

**Quadro 4 –** Palavras com freqüência igual a 1 no texto

| TAMANHO | LOCAL | FREQÜÊNCIA | PALAVRA |
|---|---|---|---|
| 8 | 542 | 61 | Palavras |
| 10 | 4 | 56 | Informação |
| 10 | 495 | 23 | Pensamento |
| 10 | 1.784 | 22 | Freqüência |
| 8 | 738 | 14 | Conteúdo |
| 9 | 251 | 14 | Narrativa |
| 12 | 106 | 13 | Conhecimento |
| 8 | 242 | 13 | Conjunto |
| 9 | 1.919 | 12 | Linguagem |
| 9 | 1.323 | 12 | Conceitos |
| 8 | 2.214 | 8 | Aparecem |
| 11 | 772 | 7 | Convergente |
| 10 | 423 | 7 | Estruturas |
| 11 | 103 | 7 | Apropriação |
| 10 | 1.242 | 7 | Cientistas |
| 8 | 97 | 7 | Receptor |
| 8 | 298 | 6 | Discurso |
| 10 | 778 | 6 | Divergente |
| 11 | 1.059 | 6 | Significado |
| 13 | 2.233 | 6 | Singularidade |

**Quadro 5 –** Palavras mais freqüentes e reveladores da relevância do texto

Quanto às palavras significativas que denotam a nuance do presente texto, aparecendo uma única vez na edição, estão elas dispostas no **Quadro 6**.

| N LETRAS | POSIÇÃO | FREQÜÊNCIA | PALAVRA |
|---|---|---|---|
| 8 | 35 | 1 | Conceito |
| 8 | 59 | 1 | Reflexão |
| 8 | 484 | 1 | Enunciar |
| 8 | 1.032 | 1 | Evocação |
| 8 | 1.254 | 1 | Explosão |
| 8 | 1.754 | 1 | Medições |
| 8 | 2.190 | 1 | Contagem |
| 8 | 2.377 | 1 | Unitermo |
| 8 | 2.910 | 1 | Programa |
| 8 | 3.493 | 1 | Conhecer |
| 8 | 3.529 | 1 | Palavra |
| 8 | 3.695 | 1 | Sensação |
| 8 | 3.699 | 1 | Educação |
| 8 | 3.815 | 1 | Inclusão |
| 9 | 665 | 1 | Cognições |
| 9 | 670 | 1 | Percepção |
| 9 | 1.033 | 1 | Simbólica |
| 9 | 1.081 | 1 | Contexto |
| 9 | 1.953 | 1 | Escritura |
| 9 | 2.375 | 1 | Holofrase |
| 9 | 2.389 | 1 | Densidade |
| 9 | 2.426 | 1 | Indexação |
| 10 | 1.144 | 1 | Morfologia |
| 10 | 1.636 | 1 | Linguagem |
| 10 | 1.775 | 1 | Científico |
| 10 | 2.741 | 1 | Semântica |
| 10 | 3.240 | 1 | Mecanismos |
| 10 | 3.532 | 1 | Ortografia |
| 10 | 3.658 | 1 | Comunidade |
| 10 | 3.664 | 1 | Computador |
| 10 | 3.710 | 1 | Totalidade |
| 11 | 440 | 1 | Fragmentada |
| 11 | 708 | 1 | Assimilação |
| 11 | 2.441 | 1 | Lingüística |
| 11 | 3.310 | 1 | Indicadores |
| 12 | 300 | 1 | Significação |
| 12 | 1.616 | 1 | Objetividade |
| 12 | 2.480 | 1 | Metodológico |
| 12 | 2.526 | 1 | Automatizada |
| 12 | 3.724 | 1 | Cibercultura |
| 12 | 3.725 | 1 | Complexidade |
| 13 | 184 | 1 | Racionalidade |
| 13 | 1.466 | 1 | Gerenciamento |
| 13 | 1.764 | 1 | Produtividade |
| 14 | 3.309 | 1 | Representação |

**Quadro 6 –** Palavras mais freqüentes e reveladores da relevância do texto, aparecendo uma única vez

Diante do exposto até então, é evidente que cada um pode fazer sua própria avaliação de como palavras deslocadas podem ser utilizadas para reconstruir o significado de uma narrativa de informação. A linguagem é universal e parte integrante da condição humana. Em oposição, a escrita / as palavras consistem em invenção relativamente moderna. Enquanto a linguagem é um sistema de signos convencionados que representam a realidade na comunicação humana, as palavras atuam como representação gráfica da língua. As mentes humanas forjadas numa existência oral não lidam bem com o jogo de palavras instituído pela escrita, até porque domar sua percepção implica rearticular cadeias de pensamento integrando novas conexões.

Por fim, afirma-se que a percepção das palavras de uma narrativa num texto escrito é mais complexa do que a sensação da cadência ancestral de uma música conhecida. A apreensão da escrita demanda decodificação, recognição e interpretação. Envolve configuração mental, a qual, por seu turno, prescinde de avaliação, memória, signos, significantes e significados. A apreensão do significado pela escrita acontece num momento presente em confluência com o passado e na perspectiva do futuro. A análise computacional das narrativas possibilita atribuir um indicador para o conteúdo, de tal modo que tal valorização seja imperceptível para a análise tanto pelo ser humano quanto pelo computador.

Um texto é um grande estoque de palavras. Palavras representam um conjunto de conceitos dentro de domínios e seus possíveis relacionamentos. A organização semântica das palavras é como uma forma de "congelar" o conhecimento existente no universo (ou parte dele) sem, contudo, substituir a narrativa original livre e completa na representação de seus enunciados.

Logo, tomando como referência o estudo das palavras retiradas no contexto do texto e da configuração de suas diferentes freqüências, é admissível fixar valores circunstanciais de relevância para uma peça de informação. Trata-se do valor de uso e da prioridade de tal uso, em determinada circunstância e para determinado receptor ou grupo de receptores.

Este é um estudo exploratório que entende que as palavras nunca estão protegidas dos riscos do desaparecimento. Frágeis, são elas fortalecidas quando em coesão entre si e com a memória que guarda o significado. Os manuscritos, assim como os poemas escritos nas árvores, se perdem ou as páginas dos livros podem se apagar. Afinal, a própria memória, também é falha!

## REFERÊNCIAS

ARAÚJO, C. A. Bibliometria: evolução histórica e questões atuais. **Em Questão: Revista da Faculdade de Biblioteconomia e Comunicação da UFRGS**, Porto Alegre, v.12, n.1, 2006.

BUCKLAND, M. Paul Otlet: pioneer of information management. **School of Information Management & Systems**. Disponível em: <http://www.sims.berkeley.edu/ ~buckland/otlet. html>. Acesso em: 8 jun. 2010.

BUSH, V. [**Sobre *As we may think*.**] Disponível em: <http://pt.wikipedia.org/ wiki/As_ We_May_Think>. Acesso em: 29 mar. 2009.

FOUCAULT, M. **O que é um autor?** 3. ed. Lisboa: Vega, 1992.

GUILFORD, J. P. Three faces of intellect. **American Psychologist**, [S. l.], v. 14, n. 8, 1959.

GUEDES, V. L. S.; BORSCHIVER, S. Bibliometria: uma ferramenta estatística para a gestão da informação e do conhecimento, em sistemas de informação, de comunicação e de avaliação científica e tecnológica. In: ENCONTRO NACIONAL DE CIÊNCIAS DA INFORMAÇÃO, 6., 2005, Salvador. **Anais...** Salvador: Universidade Federal da Bahia, 2005.

LANCASHIRE, I. Uttering and editing: computational text analysis and cognitive studies in authorship. **Texte et Informatique**, [S. l.], n. 13 / 14, p. 173-218, 1993.

ONG, W. J. **Orality and literacy**: the technologizing of the word. New York: Terence Hawkes, 1988.

SANTOS, D. **Português computacional**. Disponível em: <http://www.linguateca.pt/Diana/download/SantosCIP94.pdf>. Acesso em: 17 fev. 2010.