Línguas & Letras

ESTUDOS LINGÜÍSTICOS

ISSN: 1517-7238
vol. 7 nº 12
1º sem. 2006
p. 139-158

EXTRAÇÃO DE INFORMAÇÃO COMO UM IMPORTANTE RECURSO PARA A INTERPRETAÇÃO DE TEXTOS

PORFIRIO, Lucielen[1]
BIDARRA, Jorge[2]

---

[1] *Mestre em Letras – Programa de Pós-Graduação* Stricto Sensu *em Letras, nível de Mestrado – área de concentração em Linguagem e Sociedade da Unioeste* – Campus *de Cascavel-PR.*

[2] *Professor Adjunto do Programa de Pós-Graduação* Stricto Sensu *em Letras, nível de Mestrado – área de concentração em Linguagem e Sociedade da Unioeste* – Campus *de Cascavel-PR.*

ISSN: 1517-7238 v. 7 nº 12 1º sem. 2006 p. 139-158

Resumo: Ao tentar interpretar um texto, todo leitor vai certamente se deparar com muitos desafios. Há, sem dúvida, várias maneiras de ele tentar transpor as barreiras impostas nesse tipo de tarefa. Uma delas é, p.ex., perceber que certas palavras parecem ser mais importantes semanticamente do que outras, apesar de todas serem, no conjunto, imprescindíveis. Desenvolver uma análise lingüística tomando como base a identificação dessas palavras, aqui referidas como palavras-chaves, à luz de princípios semântico-lexicais, é o principal objetivo da pesquisa que vimos desenvolvendo já há algum tempo. Padrões Lexicais, tais como colocação, coligação e prosódia semântica, são itens com os quais trabalhamos. Partindo de um corpus no domínio da gastroenterologia e da palavra 'causa', muito freqüente em textos dessa natureza, tentaremos mostrar (i) a influência que essa palavra exerce sobre as outras com as quais se relaciona; (ii) a influência que ela própria sofre dessas outras palavras e, sobretudo, (iii) como toda essa articulação acaba afetando sobremaneira a interpretação de textos então considerados..
Palavras-chave. *Interpretação de textos, padrões lexicais, palavras-chave.*

Abstract: In the attempt of interpreting a text, any reader will face many challenges. Undoubtedly, there are many ways of trying to overcome the problems that arise during this type of task. One of them is, for instance, noticing that some words seem semantically more important than others, in spite of the fact that all of them are indispensable in group.
Developing a linguistic analysis having these words identification, here named keywords, following lexical semantic concepts is the main goal of this paper, which has been developed as a larger project for a longer time. Lexical patterns such as collocation, colligation and semantic prosody are concepts that we work with. Having a corpus on the dominium of gastroenterology and the word 'cause', very frequent in such texts, as a first bases, try to show (i) the influence that this word has on the others which it relates to; (ii) the influence it suffers from others and above it all, (iii) the way that all this articulation ends up affecting the text interpretation considered here.
Keywords: *Text interpretation, lexical patterns, keywords.*

I. INTRODUÇÃO

A interpretação de textos é, sem dúvida alguma, um tipo de processamento de alta complexidade que, para ser bem realizada, requer do leitor não apenas um conhecimento lingüístico e extralingüístico prévio, mas também um enorme esforço cognitivo. Para interpretar, todo leitor deve, no mínimo, ser capaz de decodificar o texto e, com base nisso, elaborar representações mentais que, de algum modo, contemplem as descrições sobre o que o texto quer passar como

informação. Para obter sucesso nessa tarefa, os leitores precisarão, necessariamente, levantar hipóteses, bem como fazer inferências, tendo por suporte a ativação dos conhecimentos lingüísticos e enciclopédicos que possui.

Pereira (2002) e Kleiman (2001) enfatizam que os principais conteúdos e idéias contidos num texto se expressam não só por meio dos itens lexicais, mas também e principalmente, pelas interações que as palavras estabelecem entre si. Explorar, pois, as palavras que dão corpo ao texto e analisá-las com base nas funções que desempenham no seu interior são caminhos a serem seguidos por todos aqueles que pretendem desenvolver uma boa interpretação de texto.

Vários métodos têm sido utilizados no sentido de se trabalhar com a interpretação de textos. Dentre os mais comuns, podemos citar a complementação de lacunas, a elaboração de perguntas, seguidas de respostas orais ou escritas (Colomer & Camps, 2002; Pereira, 2003), e, ganhando força mais recentemente, a extração de informação. A Extração de Informação de Textos (EIT) é um método, até certo ponto simples, porém não trivial, que consiste, basicamente, na identificação e captura de aspectos lingüísticos relevantes (lexicais, sintáticos e semântico-conceituais) contidos nos textos, tendo nas chamadas palavras-chaves a sua principal fonte de exploração.

Tomando justamente a EIT como o nosso método de trabalho, o objetivo aqui é apresentar um estudo que vimos realizando como um projeto de pesquisa. A idéia central do trabalho é investigar, bem como, ao final, tentar mostrar qual o grau de influência das palavras-chaves sobre o processo de interpretação, chamando a atenção dos leitores para esse fato.

Antes de seguirmos adiante, cabe esclarecer dois pontos importantes. Primeiramente, que as discussões que apresentaremos nesse artigo resultam de um estudo realizado a partir de um *corpus* variado, cujos textos pertencem a um domínio específico, o da gastroenterologia. Em segundo lugar, que devido à grande quantidade de palavras-chaves investigadas, apenas uma análise parcial dos dados será considerada nesta apresentação.

Para o debate, estruturamos o artigo da seguinte maneira. Na seção 2, uma discussão é realizada acerca de aspectos teóricos importantes relacionados à interpretação de tex-

ISSN: 1517-7238 v. 7 nº 12 1º sem. 2006 p. 139-158

tos e também à extração de informação. Em 3 e 4, respectivamente, passamos à descrição do método utilizado neste trabalho e à análise dos dados, propriamente dita, seguida de uma pequena reflexão a respeito dos resultados que foram obtidos até o momento nessa pesquisa. Por fim, na seção 5, apresentamos as nossas considerações finais, seguidas, em 6, pelas referências bibliográficas.

2. ASPECTOS TEÓRICOS RELEVANTES PARA A INTERPRETAÇÃO DE TEXTOS

De acordo com Colomer & Camps (2002), para que um leitor consiga interpretar um texto, ele deverá ser capaz de descobrir quais caminhos percorrer para organizar as idéias expressas no texto que está lendo. Apesar de parecer o óbvio, o fato é que nem sempre saber identificar esses caminhos é uma tarefa simples para o leitor. Isso porque, nessa sua tentativa, o sujeito, consciente ou inconscientemente, precisará levar em conta diversos elementos que não se restringem apenas aos de natureza lingüística, mas também envolvendo o seu conhecimento de mundo e, sobretudo, a sua capacidade cognitiva.

Eco (1979) e Kleiman (2001) argumentam, por exemplo, que os processos de inferenciação e levantamento de hipóteses, antes e durante a leitura, são dois itens cruciais para que o leitor seja capaz de compreender bem um texto. Para os autores, é somente testando as suas hipóteses e inferências que o indivíduo vai-se tornando apto para, por um caminho inverso ao do escritor, reconstruir os sentidos contidos no texto. Como preconiza Kleiman (2002, p. 65), "o leitor constrói e não apenas recebe um significado global para o texto. Ele procura pistas formais, antecipa essas pistas, formula e reformula hipóteses, aceita ou rejeita conclusões".

Mais exatamente, um leitor só terá condições de compreender bem o que está lendo, se, desde o primeiro contato com o texto, ele souber explorar o potencial que certas palavras adquirem dentro da obra (Kleiman, 2002; Pereira, 2003). Com base nisso, pode-se dizer que é justamente no momento em que o leitor passa a trabalhar com essas palavras, convencionalmente chamadas de "palavras-chaves", que o processo

interpretativo começa a se consolidar de fato. Admitindo, então, que as palavras-chaves constituem itens de grande relevância para a interpretação de textos, entendemos que investir num estudo mais detalhado sobre o assunto torna-se uma linha de investigação favorável aos objetivos.

Dentre as diferentes maneiras de se verificar a influência das palavras-chaves na interpretação, uma delas, segundo Sardinha (1999b), seria partindo para a verificação do comportamento dos padrões lexicais (ou regularidades lingüísticas) que elas apresentam no interior do texto. O que aqui denominamos de padrões lexicais é um conjunto de estruturas lingüísticas complexas, capazes de criar em torno de si ambientes altamente propícios ao processo interpretativo. Conforme a literatura, seriam três os principais tipos de padrões lexicais: Colocação, Coligação e Prosódia Semântica (Sardinha, 1999; Partington, 1998; Sinclair, 1991).

A colocação diz respeito a todas as palavras que podem vir associadas ou co-ocorrendo com um núcleo lexical, num mesmo sintagma. Por exemplo, a palavra 'causar' tende a ocorrer mais freqüentemente com colocados tais como 'problemas', 'prejuízo, 'danos', 'morte', 'impacto' como na sentença:

(01) O vendaval **causou** grandes *prejuízos* para a população local.

A coligação, por sua vez, se refere à companhia gramatical mantida pelo núcleo, ou seja, é o relacionamento que esse núcleo assume com palavras de determinadas classes gramaticais. P.ex., entre outras possibilidades, a palavra 'só' apresentaria uma relação de coligação com "pode+ ser + particípio do verbo principal (voz passiva)", assumindo um sentido de adversidade:

(02) Esta pesquisa **só** *pode ser concretizada* a partir da observação das normas.

Define-se como prosódia semântica um padrão lexical que, dependendo das associações feitas entre certos itens

ISSN: 1517-7238 *v. 7 nº 12 1º sem. 2006 p. 139-158*

lexicais, tem por característica conduzir a interpretação para um aspecto que tanto poderá ser positivo, negativo ou neutro, conforme a situação. A palavra 'acontecer', p.ex., tem uma prosódia semântica que será ou negativa, ou neutra, dependendo do contexto em que se insere. Um exemplo: quando acompanhada de palavras tais como 'coisa' e 'algo' (talvez 'caso'), terá uma conotação neutra:

(03) *Algo* **aconteceu** para que ela tenha mudado sua opinião.

Porém, se acompanhada de palavras como 'crime' e 'acidente', o efeito torna-se negativo:

(04) Um *crime* horrível **aconteceu** no bairro noite passada.

É apropriado comentar a importância que os subsídios fornecidos pelos três padrões para a interpretação dos textos assume. A relação entre os padrões e as informações que podem ser percebidas a partir deles, torna-se possível porque o nosso conhecimento não está relacionado a palavras isoladas ou soltas, mas fortemente pelas suas combinações e pelo conhecimento cultural enxertado nelas (Stubbs, 2001). Segundo o autor, às vezes uma palavra pode acionar esquemas ou referências a outras palavras co-relacionadas a ela, podendo-se extrair desse relacionamento a identificação do assunto do texto.

2.1 Padrões lexicais e Extração de Informação como recursos para a interpretação de textos

Há, como já mencionamos, várias técnicas que poderiam ser utilizadas para a interpretação de textos; dentre elas, a Extração de Informação (EI). Riloff (1999, p. 435) define a EI como sendo "uma forma de processamento da linguagem natural na qual certos tipos pré-definidos de informação devem ser reconhecidos e extraídos de um texto".

Embora as técnicas mais freqüentemente aplicadas à EI possam variar, no geral a opção pela análise sintática tem sido a principal abordagem dos que desenvolvem trabalhos nessa área. À guisa de ilustração, vejamos um ex. de como funcionaria o método. Seja a sentença abaixo:

(05) O parlamento foi **atacado** pelos guerrilheiros.

Assumindo que a sentença (05) faz parte de um conjunto de textos que trata do terrorismo, um dos interesses aqui poderia ser, por exemplo, descobrir quem foi o responsável pelo ataque. É evidente que qualquer um de nós, sem muita dificuldade, teria todas as condições para responder a essa pergunta satisfatoriamente. Mas a questão aqui é outra, de cunho representacional.

Os fundamentos aplicados pelo método de EI vão encontrar o suporte necessário para tal processamento nos conteúdos aprendidos nas escolas sobre a análise sintática, apenas dando a eles um caráter mais formalizado. Grosso modo, o raciocínio aplicado pela EI seria resumidamente o seguinte. Primeiramente, busca-se a identificação na sentença da palavra que vai assumir o papel de núcleo lexical (no caso do nosso exemplo, o verbo "atacar"). Feito isso, parte-se para a proposição de uma representação que explicite o relacionamento do núcleo com os seus respectivos complementos pré e pós-verbais, algo nos seguintes termos: "atacar (x,y)", para "x" e "y", respectivamente, o agente e o paciente da ação. Com base nessa representação e pelo confronto das variáveis (aqui identificadas por 'x' e 'y') com os elementos contidos na sentença, exceto o núcleo, fica agora fácil descobrir quem desempenha o quê nessa relação.

Embora correto o raciocínio, o fato é que a formulação tal como está apenas favorece um tipo de análise, particularmente voltada para sentenças em voz ativa; o que, como se vê, não é o caso do exemplo. Para solucionar o problema, o que se propõe é que, para cada situação que o exija, seja fornecida uma fórmula específica. Assim, uma representação mais apropriada para a voz passiva poderia assumir a seguinte formulação: **alvo>agente>verbo na voz passiva**.

Línguas & Letras
ISSN: 1517-7238 v. 7 nº 12 1º sem. 2006 p. 139-158

Uma vez determinada a fórmula, o que se segue nada mais é do que a execução de uma operação de casamento de padrões, nesses termos. Isolando-se o verbo (*atacado*), que na fórmula vem ao final; associando-se o sujeito sintático da sentença (*O parlamento*) com a primeira parte da fórmula, **alvo**, e a expressão seguinte ao verbo (*pelos guerrilheiros*) com **agente**, obtém-se, como se pretendia, todos os personagens da ação, e ainda, com os seus papéis bem definidos.

Mesmo que parecendo um procedimento simples, vale lembrar que no fundo, nem sempre o uso de padrões sintáticos é suficiente para, por si mesmos, permitir a extração de todas as informações relevantes do texto. Vejamos um problema típico para cuja análise o método sintático, se aplicado unicamente, falharia.

(06) O parlamento foi atacado e a organização guerrilheira diz ser a responsável

Da mesma forma, mesmo que um falante nativo da língua portuguesa, em princípio, seja capaz de concluir que a organização guerrilheira foi a autora do ataque, o que muitas vezes nós não nos damos conta é de que, para chegar a esse resultado, ele precisou executar um outro tipo de processamento que vai além do sintático. Com efeito, mais do que destrinchar as relações estruturais, não é possível avançar com as análises sem que se lance mão de, pelo menos, o módulo semântico. A abordagem adotada nesse trabalho segue justamente com esse enfoque.

3. ASPECTOS METODOLÓGICOS

Como observamos, para que seja possível interpretar um texto necessário se faz investigar como as palavras se relacionam entre si, tanto sintática, quanto semanticamente. Mas, desde que realizar um estudo desses sobre todas as palavras presentes num texto não seria nem um pouco viável, o caminho

escolhido foi analisar apenas as palavras que, de alguma maneira, fossem mais relevantes no texto. Assim, decidimos restringir a nossa discussão apenas às palavras-chaves.

Segundo Cavalcanti (1989, p. 75), uma palavra-chave constrói em torno de si uma teia de fios condutores semânticos capazes de dar informações importantes sobre o conteúdo proposicional do texto. As palavras-chaves, dadas as suas características, se mostram propensas à saliência dentro dos textos em que ocorrem, o que quer dizer que se localizam nos textos como se num plano principal, tal como um foco da descrição do tema; e ainda, compartilham um ambiente coesivo com seus colocados que, se bem explorados tanto pelos escritores quanto pelos leitores, vão permitir, a ambos, melhores condições, para a elaboração ou compreensão de um texto.

Para levar adiante as nossas análises, trabalhamos com um *corpus* de pesquisa composto por 61 textos da área da gastroenterologia, totalizando 49.088 palavras. Esses textos foram todos extraídos da internet, de autorias especializadas, porém voltados para um público leigo. Como forma de identificar as palavras-chaves, todos esses textos foram submetidos, inicialmente, a um pré-processamento, com duas finalidades principais: i) determinar as freqüências de ocorrências de cada palavra (indistintamente, palavras-chaves ou não) e ii) identificar como elas estariam distribuídas em cada um dos textos. Conforme Sinclair (1991), tais dados são importantes porque formam uma base empírica para a interpretação. Segundo o autor, é a partir deles que se descobrem as candidatas a palavras-chaves e o tipo de organização que o texto assume.

Feito esse primeiro rastreamento, o passo seguinte voltou-se mais propriamente para a seleção das palavras-chaves. Para tanto, um outro *corpus*, denominado de referência, foi arrolado. O *corpus* de referência é, por definição, um texto relativamente grande e de grande representatividade para a área (p.ex., um livro). Para a escolha desse texto, reportamo-nos a especialistas da gastroenterologia que então nos fizeram a indicação, mas que, por se tratar de uma obra extensa e pela exigüidade de espaço, não pode ser incluída nesse artigo; entretanto, segue a sua referência: Dani, 2001 - *Gastroenterologia essencial*. Assim como com o *corpus* de

ISSN: 1517-7238 v. 7 nº 12 1º sem. 2006 p. 139-158

pesquisa, também o *corpus* de referência foi submetido a um pré-processamento idêntico, com os mesmos objetivos já delineados anteriormente. Tanto para o cálculo das freqüências, quanto para a seleção das palavras-chaves, usamos a ferramenta Wordsmith Tools (Scott, 2004).

Como resultado, obtivemos 94 palavras-chaves, a que chamaremos de palavras positivas. A fim de obter uma porção mais significativa de palavras-chaves (filtragem), essas 94 palavras foram submetidas a um novo tipo de processamento, do qual resultou um conjunto de palavras, aqui referidas por palavras superchaves (Scott, 2004), mais reduzido que o anterior, porém de maior expressividade: 75 palavras.

Desse conjunto, trazemos para o debate a palavra 'causa'. À primeira vista, poder-se-ia dizer que, na medida em que essa palavra não traz em si nenhum conteúdo semântico bem determinado, pouca ou nenhuma revelação consistente poderia ser produzida a partir da sua análise. Contudo, não foi exatamente isso que constatamos. As análises feitas até agora com a palavra nos têm mostrado que ela sofre alta influência de seus colocados e se apropria da semântica dos co-ocorrentes revelando informações de alto teor interpretativo para os textos. E é exatamente esse seu comportamento, por assim dizer surpreendente, que a torna uma palavra instigante e bastante enriquecedora para a nossa discussão.

4. RESULTADOS E DISCUSSÃO

Como já mencionado, para a análise, 61 textos - o *corpus* de pesquisa - e um *corpus* de referência foram utilizados. O procedimento adotado teve como preocupação básica não apenas descobrir nesses textos a quantidade de ocorrências da palavra 'causa' (no total, 108) e seus colocados, mas também identificar os contextos lingüísticos em que a palavra aparecia. Para atingir esses objetivos, fizemos uso da ferramenta "concordanciador". A partir dela, foram gerados cenários do tipo 12 por 12; isto é, tendo 'causa' como núcleo, o que chamamos de cenário é a sua composição com 12 outras palavras à

sua esquerda e I2, à direita. Um resumo, já sistematizado e conforme a nossa necessidade, é fornecido pela tabela I abaixo.

TABELA I: LISTA DAS COLOCAÇÕES, COLIGAÇÕES E PROSÓDIA SEMÂNTICA MANTIDAS PELO ITEM 'CAUSA'

OCORRÊNCIAS DA PALAVRA 'CAUSA'		
COLOCAÇÕES	**COLIGAÇÕES**	**PROSÓDIA SEMÂNTICA PREDOMINANTE**
Dor (21) Úlcera (13) Gastrite (12) Funcional (8) Sintomas (8) Dispepsia (7) Psicológica (7) Doença (6)	Substantivos: Causa + substantivo Causa+de+ substantivo Verbo de ligação+ causa+de	Negativa

De modo geral, a palavra 'causa' tende a ocorrer no *corpus* de pesquisa tanto como verbo quanto como substantivo. No primeiro caso, há uma forte evidência de que os argumentos do verbo e os traços semânticos inclusos neles constituem-se como vinculadores de informações importantes para a identificação de quem é o causador e qual o problema (por exemplo, "*algo* causa *tal problema*"). Neste caso, o sentido assumido tem um caráter notadamente negativo sobre o experienciador. Pela observação da tabela I acima, nota-se que há um grande número de ocorrências da palavra 'causa' com colocados que indicam sintomas ('dor') ou patologias ('dispepsia', 'gastrite', 'úlcera', 'doenças'), estabelecendo para o ambiente lingüístico a expectativa da aparição da descrição patologias gástricas específicas.

Na ocorrência de 'causa' como substantivo, duas possibilidades foram destacadas: uma delas indica conseqüência e a outra, embora semanticamente tenha indicações similares, mostra-se como marcadora do tópico do texto a partir de uma coligação com verbo de ligação.

Além disso, a tabela registra também um aspecto interessante que predomina nas ocorrências de 'causa': a prosódia. O que se pôde notar em relação a isso foi o fato de que, com mais freqüência, os textos tendiam para uma interpretação negativa; ou seja, neles estariam ressaltados problemas graves de saúde gastrointestinal. Não sabemos ainda precisar

ISSN: 1517-7238 *v. 7 nº 12 1º sem. 2006 p. 139-158*

com exatidão, mas o fato é que não conseguimos identificar em nossas análises uma leitura positiva (prosódia) da palavra. Apenas como um palpite, poderíamos argumentar que isso se deu, pura e simplesmente, pela natureza dos textos analisados. Todavia, para que pudéssemos opinar de maneira mais avalizada, precisaríamos arrolar nas análises mais textos, talvez e quem sabe de fontes diferentes e que tenham características e/ou finalidades outras das que aqui foram tratadas.

Na seqüência, serão mostradas algumas possíveis distribuições que puderam ser geradas pela palavra 'causa' em diferentes contextos.

4.1 A identificação do causador de um problema gástrico a partir da palavra 'causa'.

Com base em alguns colocados da palavra 'causa' em sua realização verbal, é possível perceber a presença de um causador (agente), o qual semanticamente constitui-se como o desencadeador de um acontecimento gástrico. O ex. abaixo nos fornece alguns subsídios:

> 07) [...] levam certas *veias do esôfago e do estômago* a se dilatarem, tornando-se mais frágeis. Seu *rompimento* **causa** uma *hemorragia digestiva* das mais abundantes e difíceis de tratar [...]

No exemplo (07), as expressões 'rompimento' e 'hemorragia digestiva' estabelecem uma conotação semântica negativa que leva o leitor à identificação do problema/sintoma. Ainda, o pronome 'seu', em relação anafórica com 'veias do estômago e esôfago', fornece a informação sobre o local atingido. A partir dessas colocações, é possível construir a informação: rompimento das veias (desencadeador) = hemorragia digestiva (sintoma). Veja-se em mais um exemplo:

> 08) Na grande maioria das vezes o *soluço* **causa** não mais do que um *desconforto* com duração de poucos minutos.

Similarmente ao exemplo (07), em (08) é possível evidenciar as informações relacionadas ao causador e ao problema. A palavra 'soluço' indica imediatamente o desencadeador de alguma irregularidade no sistema digestivo, e a palavra 'desconforto' indica o sintoma por ela arrolado.

Ainda na realização da palavra 'causa' enquanto verbo, o que se observa é uma relação direta, na qual informações tais como patologias e sintomas, ficam mais evidentes através dos próprios argumentos, colocados naturais do verbo 'causar':

09) [...] é aqui no intestino delgado que a *doença de Chron* freqüentemente ***causa*** problemas de *inflamação, ulceração e estenoses*[...]

Agente/ causador (argumento 1): doença de Chron

Acontecimento/efeito (argumento 2): ulceração, inflamação, estenoses (sintomas)

Neste exemplo, identificam-se de imediato, os argumentos 1 e 2, e a ativação de seus traços semânticos propicia a identificação da patologia descrita neste trecho (Doença de Chron) e na seqüência da palavra-chave 'causa' o leitor já pode extrair os sintomas (ulceração, inflamação, estenoses).

Observe-se o fato de que a palavra 'causa' mantém em torno de si um ambiente que revela o assunto principal tratado nos textos. Com base na inter-relação dos itens lexicais e das ativações de seus significados, o leitor, aos poucos, reconstrói as informações presentes no texto. De acordo com Cavalcanti (1989), na interação de um item chave com outros itens nos textos, o leitor pode ativar esquemas e valores e estabelecer um sentido para o texto. A propriedade dos itens lexicais chaves de, por meio da associação com outros itens lexicais do texto, estabelecer sentido, revela-se com evidência nas colocações verificadas com relação à palavra 'causa'. Por si mesma, essa palavra não se mostra como portadora de informações para o assunto do texto; contudo, a sua associação com palavras indicativas de patologias ou sintomas pode

ISSN: 1517-7238 *v. 7 nº 12 1º sem. 2006 p. 139-158*

nos fornecer subsídios importantes relacionados ao conteúdo principal do texto, como é o caso do ex. (09): 'doença de Chron causa inflamação, ulceração e estenoses'.

Nota-se, também, em algumas ocorrências de sua categoria verbal, a utilização de uma pergunta para chamar a atenção do leitor sobre uma informação importante, como mostra-se em:

10) [...] o que ***causa*** o aparecimento da *úlcera*?

A utilização da palavra-chave 'causa' na pergunta faz com que o leitor antecipe algumas das informações e mantenha uma expectativa de receber essas respostas no texto. De acordo com Kleiman (2004: 111), o leitor sabe que as informações se constroem no par pergunta-resposta, e assim espera que o autor forneça essa resposta a procura durante a leitura. P.ex. a partir do ex. (10) o leitor cria uma lacuna interpretativa do tipo "_____ causa úlcera", que facilita seu trabalho interpretativo na procura de informações válidas que preencham esse espaço.

Verificou-se, também, que ocorrências tais como (10) aparecem em títulos e subtítulos dos textos de estudo. De acordo com Cavalcanti (1989), quando itens lexicais aparecem em títulos, início ou fim do texto, revelam-se como de alta relevância, pois salientam a estrutura temática do texto. Tal observação nos leva à confirmação da importância da palavra-chave 'causa' como indício interpretativo para os textos.

4.2 A indicação da conseqüência de um problema a partir da palavra 'causa'.

Além do causador de um problema, é possível ainda evidenciar na ambiência de 'causa', as conseqüências de um determinado problema gástrico no organismo. Veja-se o ex. (11):

11) [...] que uma *bactéria* chamada de *Helicobacter Pylori* é a grande *responsável* pela ***causa*** das *úlceras*. Além dela, fatores como o fumo, o estresse e medicamentos [...]

Problema (causador): presença da bactéria *Helicobacter Pylori*
Conseqüência: úlceras.

Repare-se que, no ambiente lingüístico construído pelo substantivo 'causa', informações reveladoras do problema e da conseqüência ficam claras. Por meio da identificação das palavras '*Helicobacter Pylori*' e 'responsável', o leitor obtém a informação sobre o causador do problema. A colocação com 'úlcera' revela a patologia descrita no texto. Partindo do pressuposto de que o leitor que se propõe a ler este tipo de texto procura nele informações relativas à patologia, a identificação dessas combinações não pode passar desapercebida no processo interpretativo.

Mais um indicativo da relação de conseqüência é a coligação (vide tabela I) em que aparece um verbo de ligação (verbo de ligação+causa+de), como mostra o exemplo (I2) abaixo:

I2) [...] probabilidade de desenvolver este tipo de câncer. O *câncer colo-retal* é a terceira ***causa*** mais comum de *morte* por câncer, no Brasil. Possui maior incidência [...]

Problema(causador): câncer colo-retal
Conseqüência: morte

A coligação observada funciona como um marcador do tema principal abordado no texto. Em (I2), por exemplo, a palavra 'câncer' estabelece uma carga semântica altamente negativa, reforçada pela palavra 'morte'. Em combinação, essas palavras instalam no leitor um estado alerta para a patologia e o auxiliam a retirar a informação: "câncer é causa de morte". O leitor identifica a informação importante diretamente ligada ao tópico principal a partir da situação semântica percebida em '.... é causa de....' (verbo de ligação+causa+de) e, assim, procura organizar as informações subseqüentes que possivelmente aparecerão co-relacionadas a ela no desenrolar do texto. Como afirma Kleiman (2004:94), os leitores obedecem a uma organização hierár-

ISSN: 1517-7238 *v. 7 nº 12 1º sem. 2006 p. 139-158*

quica de informações, isto é, o leitor procura perceber o tópico e a partir dele organizar todas as informações, destacando aos poucos aquelas que têm uma dependência direta com o tópico identificado. Neste caso, a coligação mantida pela palavra 'causa' parece auxiliar o leitor nessas tarefas.

Outro fato que chama a atenção é que, da mesma forma que acontecia com o verbo, o substantivo também pode aparecer ou em resposta a perguntas. O ex. abaixo nos auxilia nesta percepção:

> I3) A **causa** infecciosa. A descoberta da participação do *Helicobacter Pylori* na **causa** destas *doenças*, que hoje é tida como o agente da mais freqüente infecção humana [...]

Os itens '*helicobacter pylori*' e 'doenças' se relacionam diretamente com o título "causa infecciosa", os quais devem ser destacados pelo leitor e, na seqüência, ligados ao assunto dos textos. Ainda de acordo com Kleiman (2004), para identificar as informações mais relevantes por meio de tópicos, o leitor procura marcações formais e itens lexicais relacionados ao título.

Reitera-se que o item lexical 'causa' é vazio de conteúdo e apropria-se de traços semânticos dos seus colocados (por exemplo, 'úlcera', 'gastrite', 'doença'), naturalmente negativos. Uma relação de apropriação de sentidos, quase como a anáfora ou a catáfora fornece à 'causa' o sentido necessário para sua interpretação. Em (I3), p. ex. um sentido altamente negativo relacionado à patologia é percebido por meio da colocação com '*Helicobacter Pylori*'. Aqui, o leitor procura na palavra 'doenças' o complemento da informação exigida pelo substantivo (causa de quê?). Numa relação de catáfora, onde doenças se sobrepõe à palavra-chave 'causa', assim como '*helicobacter pylori*', o leitor é levado a reconhecer nos colocados algumas das interpretações necessárias para o texto.

Neste sentido, Kato (1999) afirma que o leitor espera que um tema ou subtema se mantenha no texto por um tempo e,

por isso, ele procura construir com as frases uma representação mental ampla. Para tanto, ele utiliza o princípio da parcimônia, que consiste em diminuir participantes, ações e eventos nessa representação. Assim, ele interpreta muitos termos como tendo uma possível relação com um antecedente. Ao identificar essa relação, o leitor integra a informação nova na estrutura da memória, ligando-a ao antecedente localizado.

5. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Ao longo do artigo, tentamos mostrar que as palavras-chaves, as relações estruturais e semânticas que estabelecem com outras palavras nos textos, bem como as suas próprias semânticas internas, referidas por nós como prosódia semântica, constituem, todas elas, informações altamente relevantes e úteis para a interpretação textual. Partindo de textos que foram produzidos especificamente no domínio da gastroenterologia e tomando a palavra 'causa' como base para as nossas análises, mostramos que, apesar de aparentemente ela ser vazia de conteúdo semântico, o modo como ela articula com as demais palavras permite que o leitor perceba e seja conduzido para diferentes possibilidades de leituras, revelando informações altamente importantes para o processo interpretativo dos textos.

Com isso em mente, partimos para a exposição do que temos feito. Embora pudéssemos nos dar por satisfeitos com os exemplos que usamos para as demonstrações pretendidas, o fato é que, no fundo, deixamos sem qualquer explicação plausível detalhes importantes que, agora, achamos por bem comentar. Com efeito, mostramos e até chegamos a afirmar que certas interpretações seriam devidas à palavra 'causa', com maior ou menor ênfase, sua prosódia semântica e seus co-ocorrentes mais imediatos, dando a entender de que se tratavam de elementos suficientes e bastantes. Porém, estamos convencidos de que para um leitor mais atento essa impressão não passou incólume.

É verdade que, ao longo de todo o debate, estivemos sempre defendendo a idéia de que o jogo estabelecido entre as palavras-chaves e seus colocados/coligados é o que, ao

ISSN: 1517-7238 *v. 7 nº 12 1º sem. 2006 p. 139-158*

final das contas, vai determinar o curso para a interpretação de um texto. Contudo, não podemos perder de vista que, nessa trama, a verdadeira interpretação não teria como acontecer de fato se não nos dispusermos a incluir também no debate a participação de outras palavras e/ou expressões que não apenas aquelas cujos papéis foram citados.

Assim, num esforço para desfazer quaisquer dúvidas que porventura possam surgir a partir do que, realmente, dissemos, vejamos através de um novo exemplo algo que possa elucidar essa nossa preocupação e para a qual devemos, sim, uma explicação mesmo que ainda incipiente. Para tanto, tomemos para a nossa referência o trecho fornecido a seguir. Usando a mesma estratégia aplicada sobre os exemplos anteriores, para os destaques, lançamos mão de marcações como o sublinhado, o itálico e o negrito para indicar, na ordem inversa, a palavra-chave (**causa**), os co-ocorrentes (*úlcera, dor* e *queimação*) e as demais palavras e/ou expressões.

(27) A *úlcera* geralmente **causa** *dor e queimação* na parte superior do abdome. Estes sintomas são mais freqüentes em jejum e aliviam com alimentação e leite. A sensação de queimação pode ocorrer na alta madrugada fazendo a pessoa acordar pela dor. Antiácidos e leite usualmente oferecem alívio temporário. Alguns pacientes sem queixa dolorosa têm fezes negras, indicando uma úlcera hemorrágica. A hemorragia é uma complicação muito séria das úlceras.

A palavra 'úlcera' conduz o leitor, de imediato, à identificação da patologia descrita. As palavras 'dor' e 'queimação' preenchem as lacunas semânticas relacionadas a sintomas e cria no leitor a expectativa para o fornecimento de outras informações relacionadas a esses assuntos, as quais podem ser percebidas na seqüência do texto através de palavras que denotam sintomas como 'dor', 'fezes negras' e 'hemorragia'.

Uma leitura mais apressada desse trecho poderia levar o leitor para um tipo de interpretação que apenas se referisse a uma descrição da úlcera. Embora essa interpretação não seja de todo descontextualizada, ela seria ainda parcial. O fato é que para se alcançar uma interpretação mais abrangente, outros elementos também presentes nessa porção precisariam ser levados em conta. Deixá-los de lado significaria para o leitor abrir lacunas importantes e decisivas que, se preenchidas, tanto poderiam confirmar as suas expectativas em relação ao entendimento do texto, ou, ao contrário, frustrá-las por completo. Ora, mas se o leitor for hábil o suficiente para ampliar o foco da sua atenção, a ponto de perceber que as ocorrências dessas expressões ou palavras são igualmente importantes para uma boa compreensão do texto, os resquícios de dúvidas quanto à interpretação logo se dissiparão.

Note-se que, para o caso do exemplo recém comentado, a leitura de que uma patologia está acontecendo apenas se confirma a partir da inserção dessas novas informações (aqui sublinhadas) que, mesmo não sendo palavras-chaves, nem colocadas ou coligadas, constituem dados complementares e cruciais para o fechamento do ciclo interpretativo.

É óbvio que precisaríamos dizer mais; ocorre, entretanto, que para surtir efeito, necessário seria estabelecer, aqui e agora, algum critério que sustente o método. Infelizmente, isso nós não conseguiremos fazer, pelo menos por enquanto. No momento, estamos justamente investindo nesse quesito, mas ainda não temos condições de apresentar resultados que possamos considerar minimamente razoáveis. Os estudos, nesse aspecto, ainda estão numa fase muito embrionária, embora já apresentando alguns avanços bastante animadores. Esperamos que em breve já tenhamos alguma resposta nesse sentido, quando então passaremos a sua publicação, com mais segurança e a fundamentação exigida para o caso.

6. REFERÊNCIAS

Cavalcanti, Marilda do Couto. *"Interação leitor-texto. Aspectos de interpretação pragmática"* Campinas: Editora da Unicamp, 1989.

ISSN: 1517-7238 *v. 7 nº 12 1º sem. 2006 p. 139-158*

Colomer, Teresa & Camps, Anna. "*Ensinar a ler, ensinar a compreender*". Porto Alegre: Artmed, 2002. Tradução de Fátima Murad. (Original publicado em 1996)

Dani, Renato. "*Gastroenterologia essencial"*. Rio de Janeiro: Guanabara Koogan, 2001. 2a. ed.

Eco, Humberto. *"Leitura do texto literário"*. Lisboa, Portugal: Editorial Presença, 1979.

Kleiman, Ângela. *Oficina de leitura: teoria e prática.* Campinas, SP: Pontes, 2001.

_____. *Texto e leitor: aspectos cognitivos da leitura*. 8ed Campinas, SP: Pontes, 2002.

Pereira, Leda Tessari Castello. Leitura de estudo: ler para aprender e estudar para aprender a ler. Campinas, SP: Alínea, 2003.

Riloff, Ellen. Information Extraction as a Stepping Stone toward Story Understanding *In: Computational models of reading and understanding* (1999): MIT

Sardinha, Tony Berber. *"Using keywords in text analysis: practical aspects*." *In: Directpapers 42*.1999a. Disponivel em: http://www2.lael.pucsp.br/direct/DirectPapers42.pdf Acesso em: 20/01/2005

_____. *"Estudo baseado em corpus da padronização lexical no português brasileiro"* Puc/SP 1999b Disponível em: http://www2.lael.pucsp.br/~tony/1999padroes_propor.pdf

Acesso em: 23/11/2004

Scott, Mike. WordSmith Tools version 4. Oxford: Oxford University Press, 2004.

Stubbs, Michael. "*Words and phrases: corpus studies of lexical semantics*". Oxford, Massachussetts: Blackwell Publishers, 2001.

Partington, Alan. *"Patterns and meanings: using corpora for English language research and teaching"* Amsterdam Philadelphia, 1998: John Benjamins

Sinclair, J. *"Corpus, concordance, collocation."* Oxford, 1991:Oup