

## A EXTRAÇÃO DE INFORMAÇÃO APLICADA À INTERPRETAÇÃO DE TEXTOS: UM ESTUDO DE CASO SOBRE TEXTOS NO DOMÍNIO DA GASTROENTEROLOGIA \*

Lucielen Porfírio (UNIOESTE)\*  
Jorge Bidarra (UNIOESTE)\*\*

**RESUMO:** Para poder interpretar um texto, todo leitor terá, necessariamente, de extrair informações importantes, nem sempre explicitadas pelas palavras que o compõem. Com base na técnica de extração de informação e também na teoria semântico lexical, desenvolve-se aqui uma análise lingüística descritiva, com o objetivo de discutir os meandros pelo qual passa a interpretação textual. Para tanto, foram utilizados princípios da lingüística de *corpus*, tendo como foco de observação as diferentes ocorrências da palavra-chave 'intestino' e suas relações com palavras co-ocorrentes, presentes em textos escritos no domínio da gastroenterologia. Com base nesse estudo, foi possível verificar que a palavra em questão, por meio de relacionamentos sintático e semântico com os seus co-ocorrentes produz informações altamente relevantes, não só para a identificação do assunto central do texto, bem como para a sua interpretação global.

Palavras-chave: Interpretação de textos; Padrões lexicais; Palavras-chave; Extração de informação.

### 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 prévio, mas também um enorme esforço cognitivo. Para interpretar, todo leitor deve, no mínimo, ser capaz de fazer uma decodificação cuidadosa do 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, então, levantar hipóteses, bem como realizar 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 a interpretação de textos. Dentre os mais comuns, citam-se a complementação de lacunas, a elaboração de perguntas, seguidas de respostas orais ou escritas (Colomer & Camps, 2002; Pereira, 2003), e, ganhando

---

\* [lucielenporfirio@terra.com.br](mailto:lucielenporfirio@terra.com.br)

\*\* [jbidarra@unioeste.br](mailto:jbidarra@unioeste.br)

força mais recentemente, a extração de informação (EIT). A E I T 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, a partir da análise das chamadas palavras-chaves. Definem-se como tal todas as palavras que no texto se revelam mais salientes e, por suas características, trazem em seu bojo uma força semântica capaz de conduzir o leitor na compreensão do texto que lê.

Tomando a EIT como o método de análise, 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, poder mostrar qual seria o grau de influência das palavras-chaves sobre o processo de interpretação textual.

Antes de avançar, cabe esclarecer dois pontos importantes. Primeiramente, que as discussões apresentadas nesse artigo resultam de um estudo realizado a partir de um *corpus* variado, a partir de textos coletados no âmbito da gastroenterologia. Em segundo lugar que, devido à grande quantidade de palavras-chaves investigadas, apenas uma análise parcial dos dados será considerada nesse momento.

Esse artigo se estrutura da seguinte maneira. Na seção 1, uma breve discussão é realizada acerca de aspectos teóricos importantes relacionados não só à interpretação de textos, mas também referentes à extração de informação, bem como ao método utilizado para análises. Na seqüência, passa-se à análise dos dados, seguida de uma pequena reflexão a respeito dos resultados obtidos até o momento. Por fim, na seção 3, são apresentadas as considerações finais e as referências bibliográficas.

## **1. Interpretação de Textos: aspectos teóricos preliminares**

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 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 e 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, investir num estudo mais detalhado sobre o assunto é uma tarefa por si só justificada e motivada.

Dentre as diferentes maneiras de se verificar a influência das palavras-chaves na interpretação, uma delas, segundo Sardinha (1999b), seria por meio da verificação do comportamento dos padrões lexicais (também referidas por regularidades lingüísticas) que elas apresentam no interior do texto. Entenda-se por padrões lexicais um conjunto de estruturas lingüísticas complexas, capazes de criar em torno de si mesmos ambientes altamente propícios ao processo interpretativo. De acordo com a literatura, seriam três os principais tipos de padrões lexicais, a saber: *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<sup>1</sup>, num mesmo sintagma. Por exemplo, a palavra ‘causar’ tende a ocorrer, mais freqüentemente, com colocados tais como ‘problemas’, ‘prejuízo’, ‘danos’, ‘morte’, ‘impacto’, como ilustrado na sentença abaixo:

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

A *coligação*, por sua vez, se refere à “companhia” gramatical mantida pelo núcleo lexical; em outras palavras, o relacionamento que esse núcleo assume com palavras de determinadas classes gramaticais. Por exemplo, entre outras possibilidades, a palavra ‘só’ tende a estabelecer 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.

Denomina-se *prosódia semântica* um padrão lexical que, dependendo das associações feitas entre (certos) itens lexicais, conduz a interpretação para o lado positivo, negativo ou neutro da mensagem passada, conforme a situação. A palavra ‘acontecer’, para citar um exemplo., quando ladeada por palavras tais como ‘coisa’ e ‘algo’ tende a assumir um papel de neutro na interpretação:

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

No entanto, quando co-ocorrendo com palavras como ‘crime’ e ‘acidente’, assume um valor negativo:

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

As relações estabelecidas entre esses padrões e as informações que trazem latentes em suas estruturas internas se revelam importantes porque, como bem aponta Stubbs (2001), a nossa capacidade de interpretar um texto não se resume à observação das palavras isoladas ou soltas, mas construída a partir das combinações que se realizam nos textos e, muitas vezes, também pelo conhecimento cultural enxertado nelas. Segundo o autor, em princípio, toda pode acionar esquemas ou referências a outras palavras que com ela se relacionam capazes de

permitir ao leitor não apenas a identificação do assunto central do texto, mas também a compreensão do texto como um todo.

### 1.1. Os padrões lexicais *versus* extração de informação: recursos para a interpretação textual

Há, como já mencionado antes, várias técnicas úteis para a interpretação de textos; uma delas, a Extração de Informação (EI). Riloff (1999, p. 435) define a EI como sendo “uma forma de processamento da linguagem natural em que certos tipos pré-definidos de informação devem ser reconhecidos e extraídos de um texto”. Embora as técnicas mais freqüentemente aplicadas no campo da EI possam variar, no geral, a opção pela análise sintática tem sido a principal abordagem dos que desenvolvem trabalhos nessa área, cujo funcionamento seria mais ou menos o seguinte. Seja a sentença fornecida abaixo:

(05) O parlamento foi **atacado** pelos guerrilheiros.<sup>2</sup>

Assumindo que essa sentença faz parte de um conjunto de textos cujo tema central é o terrorismo, uma curiosidade natural é tentar descobrir o autor ou autores do ataque. É evidente que um leitor, sem muita dificuldade, é capaz de responder a essa pergunta satisfatoriamente. Embora simples, até porque a sentença-exemplo é também muito simples, o intrigante é saber quais teriam sido as estratégias usadas por esse leitor para chegar a resposta procurada. Esse tem sido o desafio da EI.

Os fundamentos aplicados pelo método de EI, na verdade, se baseiam nos conteúdos aprendidos nas escolas, principalmente aqueles oriundos da análise sintática. O que se tenta fazer, no entanto, é dar a esses conhecimentos um caráter mais formalizado. Grosso modo, o raciocínio seria esse. 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 explicita 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, o que se segue é um trabalho de casamento de padrões.

Embora correto o raciocínio, o fato é que a formulação tal como está apenas favorece um tipo de análise, particularmente voltada para sentenças que estejam em voz ativa; o que, como se vê, não é o caso do exemplo. Para solucionar o problema, a proposta é determinar, para cada situação, uma fórmula específica. Assim, uma representação para a voz passiva seria: **alvo agente verbo na voz passiva**.

Mesmo que parecendo um procedimento trivial, vale lembrar que no fundo, nem sempre o uso de padrões sintáticos é suficiente para permitir a extração de todas as informações relevantes do texto.

(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 não se limitou ao processamento sintático. Mais do que destrinchar as relações estruturais, não seria possível avançar com as análises sem que, nesse processamento, fosse envolvido, pelo menos, o módulo semântico. A abordagem adotada nesse trabalho busca explorar esse relacionamento..

## 1.2. Da metodologia

Como enfatizado, para que seja possível interpretar um texto se faz necessário investigar como as palavras se relacionam entre si, tanto no nível sintático, quanto no semântico; uma análise geralmente feita sobre as 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. Nessa perspectiva, estão localizadas nos textos como se num plano principal, tal como um foco da descrição do tema. Compartilham um ambiente coesivo com seus colocados que, se bem explorados tanto pelos escritores quanto pelos leitores, vão permitir, a um e a outro, melhores condições para sua elaboração e compreensão.

As análises que fizemos tomaram como ponto de partida um *corpus* de pesquisa composto por 61 textos da área da gastroenterologia, totalizando 49.088 palavras. Esses textos foram extraídos da internet, tendo por critério de seleção os seguintes elementos: a) os textos terem sido escritos para um público não especializado na área; (b) apresentar uma linguagem clara e de fácil compreensão para esse tipo de leitor; (c) os textos precisariam ser assinados por um profissional especialista na área e (d) conter um número de textos possível de ser analisado linguisticamente.

Para identificar as palavras-chaves, todos os textos foram submetidos, inicialmente, a um pré-processamento, com duas finalidades principais. Uma delas, determinar as frequências de ocorrências de cada palavra (nesse momento, não se levou em conta se seriam ou não palavras-chaves). A outra, determinar como elas estariam distribuídas em cada um dos textos. Conforme Sinclair (1991), essas informações são importantes porque formam o que ele vai chamar de uma base empírica para a interpretação. Segundo o autor, é a partir disso que começariam a ser descobertas as “candidatas” a palavras-chaves e também o tipo de organização do texto.

O passo seguinte foi partir para a seleção definitiva das palavras-chaves, o que foi feito com base em comparações estatísticas. Para tanto, um outro *corpus*, denominado de referência, foi envolvido<sup>3</sup>. Para a escolha desse texto, especialistas da gastroenterologia foram consultados, com a indicação da obra de Dani (2001). Assim como com o *corpus* de pesquisa, também o *corpus* de referência foi submetido a um pré-processamento idêntico,

com os mesmos objetivos já delineados anteriormente: cálculo de frequências e distribuição das palavras nos textos.

Tanto para o cálculo das frequências, quanto para a seleção das palavras-chaves, usamos a ferramenta Wordsmith Tools (Scott, 2004)<sup>4</sup>. Como resultado, obtivemos 94 palavras-chaves, a que chamaremos de palavras positivas<sup>5</sup>. 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<sup>6</sup> (Scott, 2004), bem mais reduzido que o anterior, porém de maior expressividade. Desse filtro, resultaram 75 palavras-chaves, com maior destaque para a palavra ‘intestino’, que apareceu em 15% dos textos pesquisados (8 ao todo).

À primeira vista, poderíamos estar tentados a pensar que, na medida em que os textos analisados são da área da gastroenterologia, pouca ou nenhuma revelação consistente poderia ser produzida a partir da análise de uma palavra como “intestino”. Contudo, os resultados obtidos não confirmam isso. As análises feitas até agora nos têm mostrado que a sua influência para a interpretação dos textos vai muito além do significado literal e básico encontrado nos dicionários. É o que tentaremos mostrar na sequência.

## 2. Discussão e Análise dos dados

Como dissemos, para essa análise, trabalhamos com 61 textos - o *corpus* de pesquisa - e um *corpus* de referência. O procedimento adotado teve como preocupação básica não apenas descobrir nesses textos a quantidade de ocorrências da palavra ‘intestino’ (no total, 135) e seus colocados, mas também identificar os contextos lingüísticos em que a palavra aparecia. A fim de obter um contexto lingüístico mais consistente, procuramos trabalhar com porções textuais que nos fornecessem um universo lingüístico de co-ocorrência da palavra-chave. Portanto, optou-se por porções textuais em que a palavra “intestino” aparecia ladeada, tanto à sua esquerda, quanto à direita, por 12 outras palavras<sup>7</sup>, cujos resultados estão resumidos na tabela fornecida a seguir.

COLOCAÇÕES Palavras (qtde. ocorrências)	TIPOS de COLIGAÇÕES	PROSÓDIA SEMÂNTICA PREDOMINANTE
Delgado (53) Grosso (32) Porção (14) Duodeno (11) Parte (10) Cólon (09) Alimentos (07) Digestão (06) Estômago (06)	Substantivos pertencentes ao mesmo campo semântico de ‘intestino’ (p.ex., ‘estômago’ e ‘duodeno’).  Adjetivos que formam, junto com a palavra, substantivos compostos, tais como: ‘intestino grosso’ e ‘intestino delgado’.  Verbos que denotam direção:	Negativa, em grande parte das vezes e, com menos frequência, Neutra.  Não foi detectada (algo que talvez se explique pela natureza dos textos analisados) a presença de uma Prosódia Positiva.

	'passar' e 'continuar'.	
--	-------------------------	--

Tabela I: Padrões lexicais mantidos pela palavra 'intestino': Ambiente de Ocorrência

O que essa tabela mostra, em última instância, é que, seja por colocação ou por coligação, a palavra 'intestino' admite associações com diversas outras palavras e de classes gramaticais igualmente distintas<sup>8</sup>. Como tentaremos mostrar mais adiante, é justamente essa flexibilidade que a palavra exibe que vai permitir aos leitores extraírem dos textos as diferentes nuances interpretativas, que tanto poderão conduzi-lo para uma leitura explicativa do funcionamento do órgão, como levá-lo para questões mais diretamente relacionadas aos diferentes tratamentos clínicos envolvendo o órgão, ou ainda chamar a sua atenção para os cuidados que precisam ter com o funcionamento do órgão e de todo o sistema digestivo, do qual o intestino é uma peça fundamental.

Além disso, a tabela registra também o tipo de prosódia semântica predominante nas ocorrências da de 'intestino'. O que se pôde notar em relação a isso foi o fato de que, com mais frequência, os textos tendem para uma interpretação negativa; ou seja, neles estariam ressaltados os problemas mais graves de saúde envolvendo o órgão. Em raras ocasiões, a palavra assumia um comportamento neutro, cuja contribuição interpretativa somente era resolvida a partir de suas composições com outras palavras, cujas semânticas eram mais fortes e bem determinadas.

Entretanto, não conseguimos identificar situações em que a palavra contribuía positivamente para a interpretação. Intuímos que isso se deu em função da natureza dos textos analisados. Por enquanto, essa questão encontra-se em aberto.

## 2.1. A influência da palavra 'intestino' para a interpretação dos textos analisados

Vimos até aqui que 'intestino' é uma palavra que admite várias composições e que, conforme a ocorrência, poderá conduzir a leitura para um aspecto negativo ou neutro do texto. Daqui para frente, discutiremos como tais interferências podem ser facilmente notadas, cujos princípios podem ser explorados mais amplamente por qualquer leitor quando diante da necessidade de compreender um texto.

Uma primeira e importante constatação é que, quando 'intestino' aparece com colocados tais como 'delgado' e 'grosso', o efeito obtido é nitidamente de restrição. Pode-se afirmar que, em tais circunstâncias, o objetivo do escritor foi tão somente focar o órgão, em que a parte explicitada pelo colocado não tem outra finalidade senão a de fornecer uma localização mais precisa do órgão. Seja o exemplo seguinte.

(07) "O aparelho é introduzido através do ânus e todo o **intestino** *grosso* e a *porção* final do **intestino** *delgado* são examinados sob visão direta".

Nota-se nesse extrato que o alvo de atenção do leitor é deliberadamente deslocado para a palavra 'intestino' e para o órgão denotado; mas, não todo ele, apenas um de seus segmentos. Algo semelhante também se processa quando 'intestino' aparece com colocados como 'porção' (no mesmo exemplo), e 'parte', como mostrado na sentença que se segue.

(8) “Este ácido ajuda na digestão dos alimentos antes de entrar no duodeno (primeira parte do **intestino delgado**)”.

Ocorrências como essas têm como propósito tornar o assunto tratado pelo texto mais especializado, não permitindo que o leitor enverede em divagações ou, ainda, generalizações que, para efeitos de compreensão, podem assumir proporções desastrosas e incorretas, o que, por conseqüência, desmontaria todo o teor argumentativo e/ou explicativo construído pelo autor.

Há casos, entretanto, em que o alvo da interpretação pode ser modificado. Exemplo disso é fornecido no exemplo (9) a seguir.

(9) “...permitindo o exame detalhado do revestimento interno da porção superior do trato gastrointestinal, isto é, o *esôfago* - ... -, *estômago* e *duodeno* - primeira *porção* do **intestino delgado** -, usando um aparelho, o endoscópio...”

Nota-se agora que, enquanto a palavra ‘porção’ surge com a mesma força restritiva comentada antes, a composição de ‘intestino’ com palavras que denotam outros órgãos do sistema digestivo – esôfago, estômago e duodeno – obriga o leitor a expandir o seu campo de observação, fazendo-o considerar um sentido mais abrangente do assunto tratado, no caso, o sistema digestivo completo e não mais particularizado.

Não obstante, é ainda possível verificar situações em que a ocorrência de alguns colocados com a palavra ‘intestino’ acaba provocando um efeito diverso dos anteriormente discutidos. Em tais circunstâncias, na maioria das vezes, a tendência é que a leitura seja conduzida quase sempre para um aspecto relacionado à patologia, como podemos observar nos exemplos (10) e (11) subsequentes.

(10) “...A *diverticulite* causa *dor* na parte inferior esquerda do *abdome*, pois ali passa o *cólon sigmóide* (porção do **intestino** grosso antes do reto)...”

(11) “...Quando se fala em *úlceras*, porém, quase sempre as pessoas se referem às *úlceras pépticas*, isto é, às *úlceras gástricas* que surgem no *estômago*, às *úlceras* do *duodeno*, na junção do *estômago* com o **intestino delgado**, e mesmo às do *esôfago* que são mais raras...”

Note-se que, nesses casos, o sentido assumido por ‘intestino’ fica reduzido à função de localizador de patologias. O aparecimento das palavras ‘diverticulite’ e ‘úlceras’, em (10) e (11), respectivamente, revelam-se, semanticamente, mais fortes do que a própria palavra ‘intestino’ e, assim, conduzem o leitor para uma interpretação diferenciada. Ora, se são corretas as afirmações feitas por Cavalanti (1989, p. 75), de que as palavras colocadas com as palavras-chaves constroem fios semânticos que, ao final, vão nos permitir alcançar o significado dos textos em que ocorrem, podemos assegurar que a malha estabelecida pelo relacionamento semântico da palavra ‘intestino’ com esses colocados leva os leitores à ativação de esquemas mentais necessários ao jogo da interpretação.

Além da influência exercida pelas colocações de que já falamos, outras ainda também puderam ser identificadas. Por exemplo, quando ‘intestino’ co-ocorre com a palavra (associada<sup>9</sup>) ‘digestão’, a interpretação mais favorável aponta para um processo digestivo que se encontra em andamento ou que já foi concluído. O exemplo (12) ilustra essa situação:

(12) “... A *digestão* química é *completada* no **intestino** delgado e os *constituintes da refeição* são *absorvidos* no sangue...”

No caso de (12), a co-ocorrência das palavras/expressões em itálico com ‘intestino’ dão conta de que um processo de transformação alimentar está em curso. Observe-se que, isoladamente, ‘intestino’ assume uma participação discreta, quase neutra, porém um pouco mais marcada do que no caso anterior. Vale sempre lembrar que é exatamente esse jogo de correlação de forças estabelecida entre a palavra-chave e seus colocados que, ao final, vão determinar qual o sentido pretendido pela mensagem contida no texto.

Mas não são apenas as colocações que merecem atenção. Se de um lado elas exercem influência nas interpretações, não menos influente são as coligações. P.ex., quando a palavra ‘intestino’ se coliga com verbos, especialmente os que trazem na sua estrutura interna traços semânticos que implicam em direção ou passagem, a interpretação é imediata. Verbos como ‘passar por’ e ‘ir para’, sempre que ocorrendo com ‘intestino’ nos textos analisados, indicavam tratar-se de uma situação em que ‘intestino’ podia assumir, conforme o contexto, papéis diferenciados; ora usado como referência a um ponto de passagem, ora visto como um receptor de substâncias resultantes de processamentos alimentares oriundos do estômago, situações que serão retomadas na seqüência.

Como vimos, são várias as possibilidades interpretativas dos trechos em que a palavra ‘intestino’ ocorre. Com base nas nossas análises, p.ex., pudemos perceber que, dependendo do contexto, ‘intestino’, embora mantendo o seu significado básico, podia assumir sentidos bastante variados, criando em torno de si aquilo que aqui vamos chamar de categorias semânticas, sobre as quais falaremos na seqüência.

### 2.1.1. ‘Intestino’ na interpretação de um órgão receptor

Uma das nossas primeiras constatações foi o fato de que ‘intestino’ podia manifestar uma interpretação voltada para uma das suas funções no sistema digestivo, qual seja de um receptor. Os exemplos (13) e (14) abaixo mostram isso:

(14) “...podem sentir náuseas e ter vômitos por uma dificuldade do estômago em se esvaziar e *empurrar* os alimentos *em direção ao* **intestino** (gastroparesia)...”

Nesta situação, fica bem fácil perceber que a função do verbo ‘empurrar’ e também da expressão ‘em direção ao’ é atribuir à palavra ‘intestino’ o papel de receptor de algo provindo de outra parte do sistema digestivo. Situação semelhante se observa com a sentença (15), agora devido à locução verbal “sair para”:

(15) “...é o impecilho mecânico para as pedras que se encontram no canal da bile *saírem para* o **intestino**. Após feita a abertura da papila pode-se remover as pedras de dentro...”

Interpretação semelhante pode ainda ser obtida por tipos de construções que descrevem a trajetória dos alimentos no sistema digestivo, como mostra o exemplo abaixo:

(16) “...A *viagem continua até* o íleo, a última porção do **intestino** delgado, *chegando ao* intestino grosso...”

Veja-se que nesse caso a palavra ‘*viagem*’ e as locuções verbais ‘*continuar até*’ e ‘*chegar em*’, que trazem latentes em suas estruturas o traço semântico de movimento, são agora as responsáveis por esse sentido.

### 2.1.2. ‘Intestino’ entendido como ponto de passagem

De outra maneira, a palavra ‘intestino’ pode igualmente ser interpretada como órgão de passagem para objetos gastro-intestinais, como vemos em (17):

(18) “... das proteínas em forma de aminoácidos e das gorduras como ácidos graxos. *Depois do* **intestino** grosso, as fezes *passam pelo* reto para serem eliminadas. ...”

Aqui, a interpretação resulta do fato de que a palavra ‘intestino’ vem como preenchimento de uma informação que se refere à localização, por exigência das expressões ‘*passar por*’ e ‘*depois de*’ (passar por onde? Depois do que?). A expressão ‘*passam por*’ tem aqui o papel de preparar o leitor para a existência de um processo em curso, cuja conclusão é determinada pela ocorrência da expressão ‘*depois de*’.

### 2.1.3. ‘Intestino’ no sentido de localizador

Há situações em que ‘intestino’ aparece em co-relação com outras palavras apenas para atender a uma necessidade: localizar algo. Repare que em (19), essa sua condição fica bem evidenciada.

(19) “...A endoscopia é um procedimento que permite o exame do revestimento interno do *esôfago, estômago e duodeno* (1ª porção do **intestino** fino) ...”

O que aqui se pode ver é que ‘intestino’ aparece (“primeira porção do intestino grosso”) numa posição de menor destaque, se comparada aos seus co-ocorrentes, *esôfago, estômago e duodeno*. A sua aparição nessa sentença, na verdade, não tem outra finalidade senão a de localizar os leitores quanto aos órgãos citados. Tal leitura se confirma, obviamente, não apenas pela ocorrência dessas palavras, mas também pela expressão que lhes precede, qual seja: “exame do revestimento interno do”.

Esse sentido parece ficar mais claro a partir do exemplo (20) fornecido logo a seguir. Veja-se que aí a expressão ‘porção inicial do intestino’ aparece nitidamente como uma descrição que culmina com a localização do ‘duodeno’ no sistema digestivo:

(20) “...digestivo superior, ou seja, a boca, o esôfago, o estômago e a porção inicial do intestino, o *duodeno*. ...”

#### 2.1.4. A palavra ‘intestino’ em relação de causa e efeito.

Uma quarta possibilidade interpretativa para ‘intestino’ resulta de uma relação estabelecida entre causa e efeito. Como ilustração, vejamos o exemplo a seguir:

(21) “...*náuseas* e *vômitos*, as mais freqüentes são as que *causam irritação* do estômago ou **intestino**, como as infecções alimentares, gastroenterites agudas e medicamentos...”

Note-se que o aparecimento da expressão ‘*causam irritação*’ como colocado de ‘intestino’ atua aqui como um tipo de *prime* para que o leitor identifique uma situação que aponta para a ocorrência de algum problema sério; uma situação logo confirmada pela ocorrência das palavras ‘*náuseas*’ e ‘*vômitos*’, que representam sintomas.

Além do verbo ‘causar’, outros verbos também puderam ser identificados numa relação de causa e efeito com ‘intestino’. É o caso, p.ex., do verbo ‘perfurar’.

(22) “...o primeiro sintoma de uma úlcera péptica pode ser a presença de sangue nas fezes. Quando uma úlcera perfura o estômago ou o **intestino** delgado, a criança pode demonstrar sinais de dor...”

Enquanto essa relação de causa e efeito aflora das combinações das palavras ‘úlcera’, ‘perfura’, e ‘intestino’, propriamente, a ativação de traços semânticos negativos que podem ser percebidos nas expressões ‘sangue nas fezes’ e ‘sinais de dor’ é o que vai permitir ao leitor, finalmente, concluir que o trecho disserta sobre a existência de uma doença grave no intestino.

#### 2.1.5. ‘Intestino’ adotando um papel ativo no sistema digestivo.

Nas categorias analisadas até o momento, a palavra ‘intestino’, na maioria das vezes, apareceu como um elemento passivo. Mas, nem sempre é assim. Ela também pode assumir uma posição mais ativa, desde que como um órgão com funções específicas no sistema digestivo. Vejamos um exemplo.

(23) “...são absorvidos no sangue e nos vasos linfáticos. As funções principais do **intestino** grosso são a de *reabsorver a água* usada na digestão e a de *eliminar o alimento* não digerido e *as fibras*...”

A palavra ‘funções’ cria no leitor, no momento da leitura, uma expectativa quanto ao aparecimento das atribuições do órgão em questão, lacuna essa que vem a ser preenchida pelos verbos *reabsorver* e *eliminar*, verbos que denotam atividade. A mesma situação pode ser vista em:

(24) “Com exceção das fibras (nozes, farelo, fibras do aipo e outras fibras), seu intestino *desagrega* tudo o que você come em pequenos componentes para que o seu organismo possa usar...”

Note-se que, como no caso anterior, o verbo *desagrega* também vai atribuir à palavra ‘intestino’ essa noção de atividade, diferentemente do que vinha acontecendo com as categorias anteriores.

## 2.2. Refletindo sobre os resultados da análise

Embora as discussões anteriores forneçam indícios importantes sobre o comportamento e influência da palavra analisada (intestino) no processo interpretativo de textos na área de gastroenterologia, o fato é que 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 ‘intestino’, 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.

É verdade que, ao longo de todo o texto, defendemos a idéia de que o jogo estabelecido entre as palavras-chaves e seus colocados/coligados é o que, ao 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 se não considerada a participação de outras palavras e/ou expressões que não apenas aquelas cujos papéis já foram mencionados.

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 (**intestino**), os co-ocorrentes (*passa* e *completada*) e as demais palavras e/ou expressões.

(25) “Aqui são adicionadas substâncias que neutralizam o ácido do estômago, enzimas do pâncreas que ajudam a digerir carboidratos, proteínas e gorduras e bile do fígado que ajuda na digestão das gorduras. O alimento digerido *passa*, então, *pelos* seis metros de **intestino** delgado remanescentes; embora de menor diâmetro, o **intestino** delgado é muito mais comprido do que o **intestino** grosso ou cólon. A digestão química é *completada* no **intestino** delgado e os constituintes da refeição são absorvidos no sangue e nos vasos linfáticos....”

Uma leitura mais apressada desse trecho poderia levar o leitor para um tipo de interpretação que apenas se referisse a um deslocamento dos alimentos no interior do ‘intestino’, culminando com uma ‘digestão química’ e não muito mais do que isso. 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.

Note-se que, para o caso desse exemplo, a leitura de que um processo 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.

O mesmo se verifica com o trecho apresentado a seguir.

(26) Em um recém-nascido, o primeiro sintoma de uma úlcera péptica pode ser a presença de sangue nas fezes. Quando uma úlcera perfura o *estômago* ou o **intestino delgado**, a criança pode demonstrar sinais de dor. É provável que ela apresente febre. Em lactentes maiores e crianças pequenas, a presença de sangue nas fezes pode ser acompanhada por episódios repetidos de vômito ou dor abdominal.

Em (26), as colocações com palavras com carga semântica altamente negativa (por exemplo, ‘úlceras’) e coligações com verbos causativos (por exemplo, ‘perfurar’) prenunciam uma interpretação voltada aos efeitos da patologia em questão. A aparição de outras expressões identificadoras de sintomas, tais como ‘sinais de dor’ e ‘sangue nas fezes’, fornece ao leitor as informações necessárias para a confirmação dessa expectativa.

Tanto as colocações quanto as coligações de ‘intestino’, associadas à prosódia semântica negativa (ou em alguns casos neutra), possibilitam a extração de informações essenciais para a interpretação do conteúdo presente nos textos, tais como “De quais sintomas fala o texto?”; “Qual patologia é descrita?”; “O que acontece no processo digestivo?”

## Considerações Finais e Perspectivas Futuras

Ao longo do artigo, tentamos mostrar que as palavras-chaves, as relações estruturais e semânticas que se estabelecem no tecido textual, 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 ‘intestino’ como base para análises, mostramos que, apesar do seu aparente significado fechado (aquele obtido nos dicionários de palavras), o modo como ela articula com as demais palavras permite que o leitor perceba e seja conduzido para diferentes possibilidades interpretativas, seja pela influência que exerce sobre os outros itens, seja pela influência que recebe deles.

De acordo com Kleiman (2001, 2002) a interpretação é um processo complexo, que exige um esforço mental do leitor em interação com o autor e o texto. Para que consiga desenvolver este processo, ele deve identificar no texto os indícios informativos deixados pelo autor por meio dos itens lexicais. No caso desta pesquisa, este fato pôde ser evidenciado a partir da identificação das relações lingüísticas dos padrões da colocação, coligação e prosódia semântica, os quais auxiliam o leitor na identificação do assunto principal, levando-o a organizar mais facilmente as demais informações encontradas nos textos.

Nessa análise, foi possível confirmar que cada palavra no contexto de ocorrência da palavra-chave participa ativamente do significado da palavra-chave e, conseqüentemente, contribui para o sentido geral do texto.

No momento, estamos investindo noutros aspectos não mencionados aqui. Os estudos ainda estão numa fase bastante embrionária, embora já apresentando alguns avanços que consideramos animadores.

**ABSTRACT:** In order to interpret a text, any reader will necessarily have to extract important information, not always explained by the words that make part of it. Based on information extraction techniques and also on lexical semantic theory, on this paper, we develop a descriptive linguistic analysis with the goal of discussing the ways by which text interpretations goes by. In order to do so, corpus linguistics principles were used, focusing on the observation of the different occurrences of the word 'intestine' and its relations with the co-occurents, present on texts written at the dominium of gastroenterology. Based on the analysis, it was possible to notice that the chosen keyword, through the strong syntactic and semantic relationship its co-occurents, produces highly relevant information, not only for the identification of the main subject of the text, but also for the global interpretation of it.

Keywords: Text interpretation; Lexical Patterns; Keywords; Information Extraction.

## Notas

- <sup>1</sup> Núcleo lexical é a palavra que, por suas características lingüísticas, sejam elas sintáticas ou semânticas, se colocam como centrais para a compreensão de um texto.
- <sup>2</sup> Sentença extraída de Riloff (1999). Nossa tradução.
- <sup>3</sup> Caracterizam um *corpus* de referência: ser do mesmo domínio do *corpus* de pesquisa; ser maior do que o *corpus* de pesquisa e, ainda, ser representativo para a área (p.ex., um livro). Sardinha (1996, p. 89).
- <sup>4</sup> Ferramenta disponível em versão demonstrativa no site: [www.lexically.net/wordsmith](http://www.lexically.net/wordsmith)
- <sup>5</sup> Palavras-chaves positivas são aquelas que apresentam freqüências estatisticamente maiores no *corpus* de estudo, em comparação com o corpus de referência. (Scott, 2004).
- <sup>6</sup> Do original, Key keywords.
- <sup>7</sup> Para tanto, usamos a ferramenta *Concordancer*, disponível pelo "Wordsmith Tools"
- <sup>8</sup> As colocações foram geradas através do Wordsmith Tools e selecionadas de acordo com a ocorrência dessas com as palavras super-chaves. As coligações foram identificadas de acordo com a análise manual de cada uma das ocorrências da palavra e seu ambiente sintático.
- <sup>9</sup> Palavras associadas são também palavras-chaves que geralmente se ligam com a palavra superchave nos mesmos textos, ou seja, revelam o mesmo ambiente lingüístico de ocorrência (Scott, 2004).

## Referências bibliográficas

- CAVALCANTI, Marilda do Couto. *"Interação leitor-texto. Aspectos de interpretação pragmática"* Campinas: Editora da Unicamp, 1989.
- 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. 2<sup>a</sup>. 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. Disponível 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](http://www2.lael.pucsp.br/~tony/1999padroes_propor.pdf)  
Acesso em: 23/11/2004
- SARDINHA, Tony Berber. “*Comparing corpora with wordsmith keywords*”. São Paulo: LAEL/PUC-SP, 1996. Acesso em: nov. 2003. Disponível em: [http://www2.lael.pucsp.br/especialist/22\\_1\\_2001/BerberSardinha.pdf](http://www2.lael.pucsp.br/especialist/22_1_2001/BerberSardinha.pdf)>.
- STUBBS, Michael. “*Words and phrases: corpus studies of lexical semantics*”. Oxford, Massachusetts: 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

Ferramenta de auxílio à análise de dados:

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

Data de envio do artigo: 21 de março de 2006.