

LUCIELEN PORFIRIO

**UM ESTUDO SOBRE A RELEVÂNCIA DOS PADRÕES LEXICAIS PARA
A INTERPRETAÇÃO DE TEXTOS POR MEIO DA EXTRAÇÃO DE
INFORMAÇÃO**

UNIOESTE – Cascavel, PR

2006

LUCIELEN PORFIRIO

lucielenporfirio@terra.com.br

**Um estudo sobre a relevância dos padrões lexicais para a
interpretação de textos por meio da extração de informação**

Dissertação apresentada à banca
examinadora da Universidade Estadual
do Oeste do Paraná - Unioeste, Cascavel
- PR como exigência para obtenção
do título de mestre em Letras sob
orientação do Prof. Dr. Jorge Bidarra.

UNIOESTE – Cascavel, PR
2006

Mestranda: Lucielen Porfírio

Título da dissertação: **“Um estudo sobre a relevância dos padrões lexicais para a interpretação de textos por meio da extração de informação”**

Local: Unioeste – Universidade Estadual do Oeste do Paraná – Cascavel – PR

Data: 17 de fevereiro de 2006-01-11

BANCA EXAMINADORA

Profa. Dra. Sonia Aparecida Lopes Benites – Universidade Estadual de Maringá

Profa. Dra. Aparecida Feola Sella – Universidade Estadual do Oeste do Paraná

Prof. Dr. Jorge Bidarra – Universidade Estadual do Oeste do Paraná

Agradecimentos

A Deus, pela minha vida e por todas as graças concedidas e força para continuar sempre forte a fim de alcançar meus objetivos.

A meus pais, Miguel e Nilva Porfírio, pelo apoio e exemplo de vida e fé.

Aos meus irmãos, que sempre acreditaram na minha capacidade de desenvolver um bom trabalho, se fazendo presentes nos momentos de maior dificuldade.

Ao meu orientador, Jorge Bidarra, pelas considerações, apoio e constante preocupação em me auxiliar a desenvolver um bom trabalho.

À colega de mestrado, Clarice Corbari, pela disponibilidade e contribuições ao texto.

“Chega mais perto e contempla as palavras, cada uma delas tem mil faces secretas sob a face neutra e te pergunta, sem interesse pela resposta, pobre ou terrível que lhe deres: Trouxeste a chave?”

Carlos Drummond de Andrade, 1999

RESUMO:

A interpretação de textos é um processo complexo por natureza que depende não apenas de aspectos lingüísticos, mas também, cognitivos e extralingüísticos. Para interpretar um texto, todo leitor deve, inicialmente, ser capaz de decodificar o código desse texto e formular as representações mentais sobre o que é trazido como mensagem. Para tanto, o leitor precisará, necessariamente, levantar hipóteses, fazer inferências, e ativar seus conhecimentos prévios, tanto os lingüísticos quanto os de mundo (extralingüísticos). Além disso, o leitor deve localizar as principais idéias contidas num texto, as quais estão expressas nos itens lexicais e nas interações entre eles. Sendo assim, é razoável admitir que a identificação de termos isolados num texto e a análise das suas verdadeiras funções constituem, ambos, elementos de alta relevância para um trabalho de análise interpretativa. Vários métodos têm sido utilizados para se trabalhar a interpretação de textos. Dentre os mais comuns, citamos o exercício de respostas a perguntas (orais ou escritas) e, mais recentemente, a extração de informação. A extração de informação (EI) de textos é um método que consiste, fundamentalmente, na identificação e extração de aspectos lingüísticos relevantes (lexicais, sintáticos e semântico-conceituais), usados para diferentes tipos de finalidades, tais como: a sumarização, a categorização e a interpretação textual. A partir da localização de palavras-chaves e de estruturas lingüísticas, o objetivo do método é não só identificar, mas também extrair do texto as informações importantes que, em conjunto, possam permitir ao indivíduo compreender mais facilmente o assunto ali tratado. Assumindo que as interações entre os itens lexicais são, senão os únicos, um dos elementos mais importantes na interpretação de textos, o objetivo dessa pesquisa é discutir de que maneira essas relações poderiam ser melhor exploradas pelo leitor, para auxiliá-lo no trabalho interpretativo. Para a análise três palavras-chaves foram rastreadas num corpus de pesquisa no domínio da gastroenterologia: 'intestino', 'causa' e 'helicobacter pylori'. Com base nos padrões lexicais da colocação, coligação e prosódia semântica, as ocorrências de cada uma das palavras foram analisadas, com o intuito de verificar como as relações lingüísticas revelam sentidos e auxiliam no processo interpretativo. Como resultado, observamos que, mesmo sem acesso ao texto como um todo, a partir das ocorrências dos padrões foi possível extrair informações relativas ao assunto dos textos, bem como de aspectos importantes neles discutidos, tais como patologias, suas causas e efeitos.

Palavras-chaves: A interpretação de textos é um processo complexo por natureza que depende não apenas de aspectos lingüísticos, mas também, cognitivos e extralingüísticos. Para interpretar um texto, todo leitor deve, inicialmente, ser capaz de decodificar o código desse texto e formular as representações mentais sobre o que é trazido como mensagem. Para tanto, o leitor precisará, necessariamente, levantar hipóteses, fazer inferências, e ativar seus conhecimentos prévios, tanto os lingüísticos quanto os de mundo (extralingüísticos). Além disso, o leitor deve localizar as principais

idéias contidas num texto, as quais estão expressas nos itens lexicais e nas interações entre eles. Sendo assim, é razoável admitir que a identificação de termos isolados num texto e a análise das suas verdadeiras funções constituem, ambos, elementos de alta relevância para um trabalho de análise interpretativa. Vários métodos têm sido utilizados para se trabalhar a interpretação de textos. Dentre os mais comuns, citamos o exercício de respostas a perguntas (orais ou escritas) e, mais recentemente, a extração de informação. A extração de informação (EI) de textos é um método que consiste, fundamentalmente, na identificação e extração de aspectos lingüísticos relevantes (lexicais, sintáticos e semântico-conceituais), usados para diferentes tipos de finalidades, tais como: a sumarização, a categorização e a interpretação textual. A partir da localização de palavras-chaves e de estruturas lingüísticas, o objetivo do método é não só identificar, mas também extrair do texto as informações importantes que, em conjunto, possam permitir ao indivíduo compreender mais facilmente o assunto ali tratado. Assumindo que as interações entre os itens lexicais são, senão os únicos, um dos elementos mais importantes na interpretação de textos, o objetivo dessa pesquisa é discutir de que maneira essas relações poderiam ser melhor exploradas pelo leitor, para auxiliá-lo no trabalho interpretativo. Para a análise três palavras-chaves foram rastreadas num corpus de pesquisa no domínio da gastroenterologia: 'intestino', 'causa' e 'helicobacter pylori'. Com base nos padrões lexicais da colocação, coligação e prosódia semântica, as ocorrências de cada uma das palavras foram analisadas, com o intuito de verificar como as relações lingüísticas revelam sentidos e auxiliam no processo interpretativo. Como resultado, observamos que, mesmo sem acesso ao texto como um todo, a partir das ocorrências dos padrões foi possível extrair informações relativas ao assunto dos textos, bem como de aspectos importantes neles discutidos, tais como patologias, suas causas e efeitos.

Palavras-chaves: Interpretação de textos, extração de informação, palavra-chave, padrões lexicais.

ABSTRACT

Text interpretation is a complex process that depends not only on linguistics aspects, but also cognitive and extra linguistics. In order to interpret a text, any reader must, initially, be able to decode the language and formulate mental representations of the message brought by the text. In order to do so, he will need, necessarily, to make hypothesis and inferences, and activate his previous knowledge, either linguistics or extra linguistics. Besides, the reader must locate the main ideas of the text that are expressed in the lexical items and in the relation among them. In such case, it's reasonable to admit that the identification of isolated terms in a text and the analysis of its real function in it are both very important elements for the text interpretation work. Several methods might be used for working with text interpretation. Among the most common we have the answer to questions (oral or written) about the content of the text, and more recently the Information Extraction (IE). This one is a method that consists, fundamentally, on identification and extraction of relevant linguistic aspects (lexical, syntactic and conceptual semantic) used for different types of objectives, such as: summarization, categorization and text interpretation. Through the location of keywords and linguistics structures the method goal is identify and extract the most important information that together may allow the individual to understand the subject discussed there more easily. Assuming that the interactions among lexical items are one of the most important elements in text interpretation, the goal of this paper is to discuss in what way the reader could better explore this relation, in order to help him to interpret a text. For the analysis three keywords were tracked in a research *corpus* in the dominium of gastroenterology: 'intestine', 'cause' and 'helicobacter pylori'. Based on the lexical patterns of collocation, colligation and semantic prosody, these words were investigated, observing how the linguistic relations of each one could reveal meanings and help in interpretation process. As a result, we noticed that through the observation of the lexical patters it was possible to extract information regarding the text subject, as well as important aspects discussed in them, such as diseases, its causes, effects and treatments, even without having access to the whole texts.

Keywords: Text interpretation, information extraction, keyword, lexical patterns.

Lista de quadros e figuras

Tabela I – Lista de palavras super chave mostradas pelo programa Wordsmith Tools

Tabela II – Padrões lexicais da palavra ‘intestino’ no *corpus* de pesquisa

Tabela III - Padrões lexicais da palavra ‘causa’ no *corpus* de pesquisa

Tabela IV - Padrões lexicais da palavra ‘helicobacter pylori’ no *corpus* de pesquisa

SUMÁRIO

INTRODUÇÃO	11
1 INTERPRETAÇÃO DE TEXTOS: ASPECTOS TEÓRICOS PRELIMINARES:	10
1.1 A Linguística de <i>corpus</i>	26
1.2 Padrões Lexicais: O que são?.....	27
2 TÉCNICAS APLICADAS NA INTERPRETAÇÃO DE TEXTOS.....	31
2.1 Extração de Informação.....	36
3 SOBRE O MÉTODO.....	42
3.1 A seleção do <i>corpus</i> de pesquisa.....	42
3.2 O <i>corpus</i> de referência.....	43
3.3 O papel das palavras chaves na interpretação de textos.....	44
3.3.1 A seleção das palavras-chaves.....	46
4 UMA ANÁLISE APLICADA À INTERPRETAÇÃO DE TEXTOS.....	51
4.1 As relações contextuais da palavra intestino.....	52
4.1.1 'Intestino' como locativo-receptor de objetos gastro-intestinais.....	55
4.1.2 'Intestino' como ponto de passagem dos objetos gastro-intestinais.....	57
4.1.3 A palavra 'intestino' como localizador de outros órgãos.....	58
4.1.4 A palavra 'intestino' e a relação de causa e efeito.....	59
4.1.5 A ocorrência de 'intestino' como órgão com papel específico e ativo no sistema digestivo.....	61
4.1.6 Extração de informação interpretativa a partir dos padrões lexicais da palavra 'intestino'.....	62
4.2 As relações contextuais da palavra 'causa'.....	65
4.2.1 Relação causal: determinante de um acontecimento.....	66
4.2.2 A palavra 'causa' como indicadora de uma relação de consequência.....	70
4.2.3 Extração de informação interpretativa a partir dos padrões lexicais da palavra 'causa'.....	74
4.3 As relações contextuais da palavra 'Helicobacter Pylori'.....	77
4.3.1 A relação conceitual no ambiente lingüístico de 'HP'.....	80
4.3.2 A indicação da palavra 'HP' como causador de patologias.....	81
4.3.3 A relação lingüística entre 'HP' e uma necessidade de eliminação.....	82
4.3.4 A relação de 'HP' com indicadores de pesquisas.....	84
4.3.5 Extração de informação interpretativa a partir dos padrões lexicais da palavra 'Helicobacter Pylori'.....	85
CONSIDERAÇÕES FINAIS.....	88
REFERÊNCIAS BIBLIOGRAFICAS.....	92
ANEXOS	
ANEXO A – Lista das Concordâncias das palavras-chaves (as palavras-chaves retiradas do corpus em uma janela de 12:12).....	96
ANEXO B - Tabela completa das palavras super-chave fornecidas pelo programa wordsmith tools.....	106

INTRODUÇÃO

Interpretar um texto significa produzir sentidos com base em um código lingüístico, ou seja, compreender e (re)formular as idéias que estão expostas. Para que essa atividade seja realizada satisfatoriamente, requer-se o desenvolvimento de um trabalho cognitivo complexo, abrangendo aspectos lingüísticos, textuais, conhecimento prévio, entre outros.

O processo de interpretação de texto depende de uma intensa interação entre leitor, autor e texto (KLEIMAN, 2002). No que diz respeito ao leitor, é da sua competência decodificar o texto, localizar as pistas deixadas pelo autor e formular representações mentais sobre as informações ali contidas. Nesse processo, o leitor levanta hipóteses, faz inferências, ativa seus conhecimentos prévios e os aplica sobre o texto, na tentativa de o compreender. Por meio da identificação dos itens lexicais selecionados pelo autor e da interação entre eles, o leitor deve reconhecer quais são as idéias reveladas no texto. Assim, identificar termos isolados e analisar as suas verdadeiras funções no interior do texto constituem atividades de grande importância para o processo interpretativo, uma vez que as palavras contribuem, por meio de suas propriedades intrínsecas, para o significado das unidades mais complexas, como as sentenças, por exemplo (CRUSE, 1987).

O processo de interpretação necessita de aprendizagem e organização por parte do leitor. Para auxiliá-lo nesse processo, várias técnicas podem ser utilizadas, tais como a complementação de lacunas com possíveis palavras, a busca de respostas a perguntas (orais ou escritas) nas fases de pré e pós-leitura, e, ainda, a identificação de palavras-chaves do texto e a extração de informação.

A extração (automática) de informação (EI) de textos, técnica de interesse nesse trabalho, é uma área de investigação relativamente nova, mas que vem atraindo um significativo número de pesquisadores de diversas áreas do conhecimento, em particular os chamados lingüistas computacionais. O processo de EI consiste, fundamentalmente, na identificação e extração de aspectos lingüísticos relevantes, sejam eles de natureza lexical, sintática ou semântico-conceitual, permitindo a realização de diferentes tarefas, dentre as quais a sumarização do texto e a sua própria interpretação (RILOFF, 1999). Em outras palavras, com base em critérios lingüísticos, as informações relevantes são retiradas de um conjunto de textos de um determinado domínio e organizadas de acordo com o objetivo do sistema. Para que essas informações sejam identificadas e extraídas, a EI utiliza as palavras-chaves. A partir delas e de estruturas (geralmente sintáticas) pré-definidas como padrões de extração, o sistema de EI torna-se capaz de identificar e extrair do texto o assunto abordado, bem como as informações mais importantes relativas a esse assunto. A observação da ocorrência e da ambiência de tais palavras torna-se, assim, essencial nesse processo.

Na perspectiva acima delineada, **segundo uma abordagem multidisciplinar que abrange princípios relativos à interpretação de textos propriamente dita, lingüística, extração de informação e um *corpus* no domínio da gastroenterologia**, o objetivo deste trabalho se concentra em observar qual é a interação dos itens lexicais entre si em um texto, e de que maneira essa relação pode influenciar a interpretação. Para tanto, **do *corpus* selecionado**, rastream-se algumas palavras-chaves com a finalidade de identificar as influências semântico-lexicais das mesmas na extração de informações relevantes ao processo de interpretação dos textos.

Como base para a análise, foram considerados os padrões lexicais da colocação, da coligação e da prosódia semântica (SARDINHA, 1999a, 2000a; PARTINGTON, 1998; SINCLAIR, 1991), os quais permitem identificar, dentre outros elementos, o conteúdo que está sendo tratado no texto. Esses padrões foram analisados em cada uma das palavras-chaves selecionadas para o estudo, com o objetivo de investigar quais informações poderiam ser extraídas.

Para a apresentação desse trabalho, o texto foi organizado em cinco capítulos. O capítulo 1 apresenta a fundamentação teórica referente ao processo de interpretação de textos, discorrendo sobre como tal processo geralmente ocorre. O capítulo 2 contém discussões relacionadas aos tipos de técnicas mais utilizadas para a interpretação de textos e à maneira como elas podem auxiliar tal processo. No capítulo 3, relatam-se informações sobre os procedimentos adotados para a seleção dos textos e para a escolha das palavras-chaves. No capítulo 4, os dados são expostos e analisados, tendo em vista o contexto de ocorrência de cada palavra e os padrões lexicais de que ela faz parte. Por fim, nas considerações finais, apresenta-se uma descrição geral dos resultados obtidos nesta pesquisa.

1 INTERPRETAÇÃO DE TEXTOS: ASPECTOS TEÓRICOS PRELIMINARES

O processo de interpretar um texto vai muito além da mera leitura: implica saber, dentre outros aspectos, que caminhos percorrer para organizar as idéias que nele estão contidas (PEREIRA, 2003). Contudo, saber identificar esses caminhos não é uma tarefa fácil, pois envolve vários aspectos, complexos por natureza, que incluem questões cognitivas, lingüísticas e extralingüísticas.

A compreensão de um texto representa uma possibilidade de desenvolvimento intelectual, de representação e formulação do conhecimento que todo leitor está sujeito a experimentar. Porém, uma questão que vem intrigando pesquisadores é saber que mistérios subjazem a esse processamento mental dos indivíduos que os habilita a realizar a tarefa de uma maneira eficaz. Quando um leitor consegue interpretar e compreender as idéias contidas num texto, o exercício mental a que ele se submete envolve múltiplas possibilidades e esquemas, e requer aprendizado. Embora não se tenha certeza ainda sobre tudo o que ocorre no nível cognitivo do indivíduo durante a leitura, evidências apontam que são vários os mecanismos desencadeados no processo, sobre os quais, provavelmente, nem mesmo o leitor tenha consciência.

Várias abordagens têm sido adotadas para explicar o processo de interpretação de textos. Algumas das principais idéias são, resumidamente, expostas a seguir. Para a Análise do Discurso (AD), por exemplo, interpretar significa explicitar como o texto produz diferentes sentidos (ORLANDI, 1996). Na perspectiva da AD, o texto só pode ser interpretado com base em um arquivo mental já formado culturalmente pelo leitor e em uma memória discursiva. Segundo a autora:

a compreensão é a apreensão das várias possibilidades de um texto. Para compreender, o leitor deve se relacionar com os diferentes processos de significação que acontecem no texto. Esses processos, por sua vez, são função da historicidade, ou seja, da história do(s) sujeito(s) e do(s) sentido(s) do texto enquanto discurso. (ORLANDI, 1996:56).

Para Antonini (2003), a interpretação é um ato social e político inserido em um momento histórico, de acordo com o qual o leitor identifica e produz sentidos. Segundo a autora, interpretar um texto significa gerar ou formalizar conhecimentos sobre algum aspecto, colocando em evidência o significado intencionado pelo autor. Evidencia-se, então, nesse processo, não apenas o papel do leitor, mas também o do autor. Allende & Condemarin (1987) lembram que a interpretação de textos é uma atividade que depende tanto de quem os produz quanto de quem os lê. Quando se propõe a escrever um texto, todo autor tem uma determinada intenção e adota suas próprias estratégias para que o leitor compreenda e alcance aquilo que ele escreveu. O leitor, por sua vez, ao se deparar com um texto, tem diante de si uma quantidade de novas idéias, que precisam ser decodificadas e compreendidas.

É consenso entre os estudiosos que a interpretação adequada das idéias veiculadas por um texto requer do leitor um trabalho cognitivo bastante intenso, em que duas etapas, no mínimo, estão sempre presentes: a decodificação e a compreensão. A decodificação consiste na identificação de um signo gráfico, ou seja, na decifração do código lingüístico utilizado pelo escritor. Nesta etapa, os sentidos podem não estar desvelados ainda, pois o leitor pode decifrar o código, mas não ter a devida compreensão da idéia presente no texto. Entretanto, para que o leitor possa desenvolver uma compreensão do texto, os signos devem se apresentar dotados de sentidos. Dessa forma, seria somente na etapa da compreensão que o leitor começaria

a montar imagens mentais sobre o conteúdo do texto por meio do levantamento de hipóteses, de inferências e da busca dos significados que vão produzir um sentido completo para o texto lido (ALLIENDE & CONDEMARIN, 1987).

As etapas de entendimento de um texto podem abranger ainda outros aspectos, de certa forma atrelados a esses. Para Eco (1979), a interpretação de textos é resultado de um processo que depende, necessariamente, da cooperação do leitor. De acordo com o autor, que faz uma discussão baseada em textos narrativos, esse processo se desenvolve em duas ramificações principais: a das intensões e das extensões. No âmbito das intensões, estariam as informações como a enciclopédia mental do leitor, o conhecimento do código lingüístico, o saber do indivíduo a respeito da tipologia e organização dos textos (as estruturas discursiva e narrativa), as estruturas actanciais, isto é, os papéis ocupados pelos personagens presentes no texto, e, ainda, as estruturas ideológicas (julgamentos de valor, tais como a noção de bom e mau). No outro ramo da bifurcação, encontrar-se-iam outros tipos de informação, incluindo aspectos tais como as circunstâncias e o momento em que o texto foi produzido, a formulação de hipóteses e as inferências primárias e, ainda, o conhecimento de mundo, que leitor usaria mais tarde para a reformulação do texto.

Embora Eco apresente seu modelo de processamento como uma composição de vários submódulos (figura 1), sugerindo um funcionamento seqüencial consecutivo, ele observa que não há caminhos obrigatórios nessa ordem de execução. De acordo com o autor, a ordem de execução depende, fundamentalmente, da proficiência do leitor.

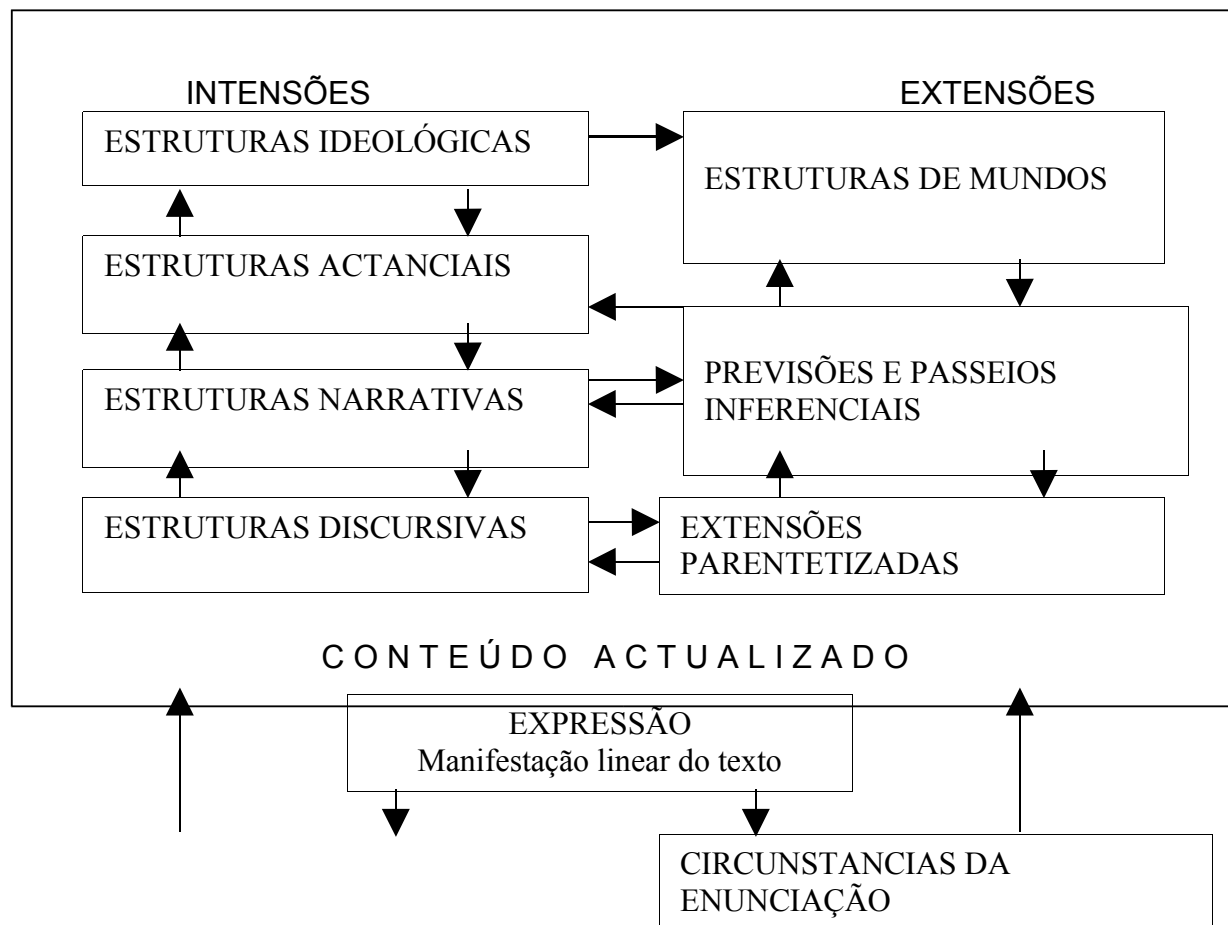


Figura 1: Os Níveis de cooperação textual.¹

A figura (1) demonstra o funcionamento do sistema de cooperação. Avaliando o diagrama de baixo para cima, verificamos que o primeiro movimento do leitor é tomar o texto como objeto de leitura (manifestação linear do texto) e aplicar sobre ele todos os seus “códigos e subcódigos”². Estes são obtidos paralelamente a partir de acessos feitos à sua enciclopédia mental por meio de seleções de contexto e da combinação de

¹ In ECO (1979: 76).

² Segundo Eco (1979), os códigos e subcódigos abrangem o conhecimento lingüístico do leitor aplicado ao texto, como, por exemplo, o acesso ao léxico e a localização de propriedades semânticas das palavras.

quadros textuais (*frames* ou esquemas), a fim de transformar o texto em um nível primário de conteúdo. Paralelamente, ele também promove a ativação das estruturas do texto que está sendo lido, do tipo de enunciado, da linguagem utilizada, para, com base nisso tudo, fazer suposições sobre o contexto social do texto. Eco denomina essa fase do processamento “circunstâncias da enunciação”. Todavia, ressalta-se que essas etapas, por assim dizer, básicas de execução não são independentes do restante do processamento. Para o funcionamento do sistema, deve haver uma dinâmica articulação com os níveis superiores do diagrama, representados, de um lado, pelo ramo das intensões, e de outro, pelo das extensões.

Ao aplicar o “código lingüístico” e as “circunstâncias da enunciação” ao texto, o leitor encontra indícios que lhe permitem fazer inferências primárias: as “extensões parentetizadas”, que são indícios inferenciais ativados pelo leitor, mas deixados em suspenso até que o texto seja melhor atualizado; as inferências são checadas conforme o desdobramento da leitura. Em outras palavras, o leitor espera encontrar os indícios necessários para confirmar as informações iniciais percebidas no material escrito; caso contrário, ele tem de reformulá-las de acordo com as informações do texto.

Mediado pelas inferências primárias, o leitor tenta, a partir daí, encontrar uma explicitação semântica de algumas partes do texto que ele já sabe serem essenciais para a sua compreensão. Para tanto, ele vai se basear na “estrutura discursiva” e nos traços semânticos fornecidos pelos itens lexicais, a fim de fazer o reconhecimento do tema (*topic*) do texto. É no módulo das estruturas discursivas, portanto, que o assunto do texto é identificado (ECO, 1979). Esse reconhecimento não se dá de uma única vez ou em bloco. Ele acontece em partes e conforme as hipóteses primárias do leitor. Com

base nas partes representadas mentalmente pelo leitor a partir do assunto e dos itens lexicais já reconhecidos, ele consegue, enfim, ampliar as propriedades semânticas do texto e estabelecer um nível de coerência interpretativa (isotopia), ou seja, “um conjunto de categorias semânticas redundantes que tornam possível a leitura uniforme de uma história” (ECO, 1979:97). Conforme o autor, “o reconhecimento do tópico é um movimento cooperativo (pragmático) que orienta o leitor para o reconhecimento das isotopias como propriedades semânticas de um texto.” (ibid.: 107).

Para que o leitor logre êxito quanto ao reconhecimento das partes semânticas de um texto, necessário se faz que ele ative os itens lexicais com os quais se depara durante a leitura e que ele próprio considera como sendo os mais importantes. A representação desses itens lexicais, segundo Eco, constituem estruturas mentais do tipo *frames*, que são quadros de representações ativados pelo leitor durante todo o processo de compreensão de um texto. À medida que a leitura evolui, esses quadros se reorganizam e ocorre um processo de formulação do pensamento. O modo de organização das informações a partir de *frames* constitui uma característica fundamental para o processo de interpretação de textos.

A partir desses esquemas de representação, o leitor deve, na etapa seguinte, ser capaz de sintetizar em uma única estrutura todos os quadros obtidos até o momento, a fim de produzir uma “estrutura da narrativa” da estória. Nesse módulo, denominado por Eco “macroposição” de esquemas, o leitor identifica uma seqüência de ações percorridas durante o texto e as “encaixa” para formular a representação da narração da história.

Se, no entanto, durante a evolução da leitura, o leitor percebe algo no texto que seja contrário à hipótese que já formulou e que, de alguma maneira, provoque a

“quebra” da representação que fez, ele se vê forçado a estabelecer novas expectativas e possibilidades e a fazer disjunções, baseado em seu próprio conhecimento, promovendo novas inferências. Trata-se, segundo Eco, das previsões e dos passeios inferenciais. Ao identificar uma nova ação que não era esperada em sua análise, o leitor é levado a fazer uma previsão sobre qual será o novo curso dos acontecimentos. Para isso, ele tem de “sair do texto” e fazer passeios sobre as possíveis inferências e, baseando-se em seu conhecimento de mundo, estabelecer os chamados “mundos possíveis” para a história. Nessa revisão, o leitor estabelece um mundo cultural com um certo número de personagens, que possuem determinadas propriedades e que assumem determinadas atitudes de acordo com o texto, ao mesmo tempo em que se apóia em sua própria cultura, confrontando esses “mundos possíveis” com o seu “mundo real”. Isso é realizado no módulo das “estruturas de mundos”, fase em que o leitor faz julgamentos de valor-verdade que estariam formados no seu léxico mental e no contexto em que vive.

Por meio das inferências realizadas com base na narrativa, o leitor passa a construir algumas lacunas abstratas ocupadas pelos “participantes” do texto e por suas respectivas características e funções – as “estruturas actanciais”, na denominação de Eco. O estabelecimento de actantes guia as opções, previsões e macroposições de quadros feitas pelo leitor no decorrer da leitura. A formulação e o preenchimento dessas lacunas são essenciais no processo interpretativo, porque indicam que o leitor foi capaz de reconstruir as informações do texto adequadamente.

Sobre estas lacunas formadas a partir do texto, o leitor aplica seus próprios julgamentos, formados por “estruturas ideológicas” presentes no seu código enciclopédico, as quais se manifestam por juízos de valor (tais como a noção de

verdadeiro e falso). Assim, o leitor confere às lacunas actanciais marcas ideológicas, que podem ser tanto negativas quanto positivas, presentes no seu conhecimento, e é induzido a conferir perspectivas ideológicas ao texto.

Apesar da análise de Eco estar baseada em textos narrativos, há elementos considerados por ele que são válidos para qualquer processo relativo à interpretação.

Para o autor, uma vez que os níveis não ocorrem de maneira hierárquica, mas concomitantemente, todas as estruturas de representações semânticas formuladas tornam-se mais claras conforme a evolução da leitura, ainda que nem sempre de forma completa. Um leitor proficiente não decodifica, mas percebe as palavras globalmente, guiado pelo seu conhecimento prévio e por suas hipóteses de leitura. Durante o processo, ele faz inferências e levanta hipóteses sobre significados que não estão diretamente presentes no texto e os incorpora ao seu conhecimento.

Se corretas as postulações de Eco, é fácil perceber que o processo de inferenciação e levantamento de hipóteses é essencial para a interpretação, porque ao testá-las em confronto com o texto, o leitor depreende o conteúdo e reconstrói sentidos textuais e, conseqüentemente, o seu próprio conhecimento. Quando o objetivo da leitura está claramente definido, as hipóteses são formuladas previamente pelo leitor e a compreensão é facilitada, pois ele ativa e percebe mais rapidamente as informações presentes no texto através da ativação mental dos itens lexicais e da percepção da relação entre eles.

Nessa mesma linha de raciocínio, Colomer & Camps (2002:23) salientam que “o leitor baseia-se em seus conhecimentos para interpretar o texto, para extrair um significado e esse novo significado, por sua vez, permite-lhe criar, modificar, elaborar e incorporar novos conhecimentos em seus esquemas mentais”. Em outras palavras,

nota-se tanto em Eco quanto em Colomer & Camps, que o texto escrito propicia converter as interpretações em realidade de conhecimento articulado, o que ocorre numa interação intensa entre pensamento e linguagem. Durante esse processo, o leitor não registra tudo o que lê de forma uniforme; pelo contrário, ele só registra aquilo que julga necessário. Tudo o mais permanece semanticamente implícito. Desse modo, o leitor deve perceber as relações entre os itens registrados durante a leitura e formular a representação do texto. Por isto, a identificação de termos isolados, assim como uma análise semântica criteriosa dos itens lexicais e da sua ambiência de ocorrência, tornam-se necessários para a eficaz compreensão de um texto. Voltaremos a falar mais sobre isso no decorrer deste trabalho.

Para Kleiman (2002), a leitura compreensiva depende da interação à distância entre autor e leitor através do texto, que se consolida por meio dos indícios deixados nos textos pelo autor sobre suas intenções, opiniões, etc., e que o leitor deverá identificar para, a partir daí, tentar reconstruir o significado que estaria presente no texto.

É provável que o leitor, para a compreensão do que está lendo, precise lançar mão de itens formais e, a partir deles, estabelecer ligações que o auxiliam nesse trabalho. De acordo com Kleiman (2002: 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”. Para a autora, a partir do conhecimento já elaborado, o leitor se vê em condições de transformar as informações presentes no texto em um novo conhecimento. Para isso, entretanto, necessário se faz que o leitor aprenda a dominar as estratégias de controle do ato de interpretação,

percebendo as marcas formais do texto e estabelecendo relações entre as idéias deste texto.

Ainda de acordo com Kleiman (2002), o leitor constrói sentidos estabelecidos no texto através de um diálogo mental com o autor. Para tanto, o leitor utiliza seus conhecimentos prévios, isto é, seus conhecimentos lingüísticos, textuais, e extralingüísticos (conhecimento de mundo) que, de todas as maneiras, contribuem para a interpretação. O conhecimento lingüístico a que se refere Kleiman (2002) é o processo no qual o leitor identifica as palavras, ou unidades lingüísticas, e as agrupa em frases (ou unidades maiores) para, em seguida, ativar significados com base em seu conhecimento sintático, lexical e semântico. Sem o domínio do conhecimento lingüístico, a compreensão do texto torna-se impossível.

Já o conhecimento textual constitui o “conjunto de noções e conceitos sobre o texto”, ou seja, é a identificação do tipo de texto e da forma de discurso que o leitor está lendo. Quando o leitor tem alguma proficiência, ele já dispõe de uma estrutura formada do texto informativo, descritivo, expositivo, argumentativo, etc., e ativa essa estrutura quando se depara com cada tipo de texto. Quanto maior for seu conhecimento textual e contato com os diferentes tipos de texto, mais fácil será sua compreensão (KLEIMAN, 2002).

Em qualquer trabalho de interpretação, como já dissemos anteriormente, o leitor necessita, ainda, aplicar conhecimentos já adquiridos e mentalizados. Este tipo de conhecimento está estruturado mentalmente em situações, as quais, para serem verbalizadas ou escritas, precisam ser ativadas. Ao escrever o texto, o autor ativa esse conhecimento de mundo e utiliza determinado conjunto de itens lexicais para expressar suas idéias; o leitor, por sua vez, ao deparar-se com estes itens, deve ativar suas

próprias estruturas mentais, seu conhecimento de mundo e as representações semânticas destes itens para compreender as informações do texto.

Todos estes tipos de conhecimentos se articulam no processo interpretativo, conforme Kleiman (2002:26) destaca:

O conhecimento lingüístico, o conhecimento textual, o conhecimento de mundo devem ser ativados durante a leitura para poder chegar ao momento da compreensão, momento esse que passa despercebido, em que as partes discretas se juntam para fazer um significado.

Segundo Pereira (2003), as marcas lingüísticas ativam os procedimentos intelectuais do leitor. Tais marcas devem ser observadas para que haja a compreensão do que se está lendo, pois é através delas que o leitor seleciona, compara e faz inferências. Em seu estudo, a autora apresenta algumas técnicas de auxílio à interpretação de textos, partindo da idéia de que a boa interpretação depende da localização dos indícios deixados pelo autor e da coleta das informações para construir o sentido do texto. À medida que o leitor percebe as palavras e estabelece para elas conteúdo semântico, sua mente vai construindo e reconstruindo significados gerais, recriando, percebendo e avaliando o texto.

Vale enfatizar a importância da representação dos itens lexicais e do conhecimento de mundo do leitor como fatores fundamentais no processo interpretativo. Contudo, deve-se ter o cuidado de perceber que não é apenas a soma dos itens lexicais isolados e da sua representação em esquemas que levará o leitor a uma construção do significado de um texto, mas a identificação das relações estabelecidas entre as palavras, de um lado, e as sentenças, parágrafos e o próprio texto, de outro. As palavras não se relacionam apenas estruturalmente, mas também

semanticamente. Conforme Colomer & Camps (2002:129), “as relações semânticas entre as palavras do texto é uma das bases de compreensão da coerência de um texto e um elemento importante para dar-lhe coesão”. Isto quer dizer que as relações semânticas entre os itens lexicais levam o leitor a estabelecer conexões essenciais para o processo interpretativo.

Cruse (1987) afirma que as palavras apresentam significado devido às suas relações contextuais, que, em última instância, são as relações de normalidade mantidas pelos itens lexicais em contextos gramaticalmente apropriados. Para o autor, o significado particular de uma palavra participa ativamente do significado de outra dentro do mesmo contexto. O que tanto Cruse quanto Colomer & Camps querem dizer é que pequenas unidades das frases de um texto se sobrepõem formando novos significados. Uma unidade (por exemplo, as palavras) é dependente da outra para a boa formação semântica da oração e, conseqüentemente, do texto.

Cruse reforça, ainda, que os itens lexicais refletem-se por suas propriedades semânticas em contextos reais e potenciais. As características combinatórias das palavras nos enunciados dependem tanto das propriedades semânticas, quanto das sintáticas. Cada palavra contribui, por meio de seus traços semânticos específicos, para o significado de unidades maiores, tais como as sentenças. Com efeito, em uma situação sentencial tal como “a_____ *está ruminando no pasto*”, o arranjo se estabelece de tal maneira que a palavra mais indicada para preencher a lacuna seja ‘vaca’ ou alguma outra do gênero. Esse determinismo, por assim dizer, se sustenta na medida em que “ruminar” e “viver no pasto” são traços semânticos presentes de algum modo na descrição do item lexical “vaca”.

Uma vez que as palavras de uma frase mantêm relações estreitas entre si, é necessário perceber como essa relação acontece. Na lingüística, uma das áreas preocupadas em estudar a relação das palavras na língua em uso é a Lingüística de *corpus*. Por meio da observação de textos em língua natural, os pesquisadores desta área procuram apontar, entre outros aspectos, as maneiras pelas quais as palavras estão interligadas produzindo significados.

1.1 A Lingüística de *corpus*

De acordo com Sardinha (2000d), a Lingüística de *Corpus* (LC) é uma sub-área da lingüística que se ocupa da coleta e da exploração de *corpora*, isto é, de um conjunto de dados lingüísticos textuais coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística.

Os estudos feitos em LC se desenvolvem segundo duas abordagens básicas: uma empirista e outra da linguagem como sistema probabilístico (SARDINHA, 2000d). O empirismo, neste contexto, significa dar prioridade aos dados oriundos da observação da linguagem, reunidos em um *corpus*. A segunda abordagem, por sua vez, baseia-se na descrição da língua a partir da probabilidade de ocorrência de traços lingüísticos em um texto. De acordo com Sardinha (2000d: 351), o estudo das frequências desses traços possibilita a percepção de que a linguagem é padronizada, pois “forma padrões que apresentam regularidade e variação sistemática (correlacionam-se com variedades textuais, genéricas, dialetais, etc.)” .

Para Stubbs (2001), uma das preocupações da LC é descrever o que é usual e típico na língua em uso, tarefa que pode ser realizada por meio da identificação dos

padrões lexicais. Esta é, de certa maneira, a principal tarefa a ser executada, se não a mais importante, pois ajuda a revelar significados.

De acordo com Sinclair (1991), é a partir da observação de uma palavra que se torna possível fazer previsões sobre o ambiente textual no qual ela provavelmente ocorre, bem como fazer previsões sobre as palavras possíveis de ocorrerem nesse ambiente, dois aspectos fundamentais para a interpretação de textos.

Pela observação do comportamento lingüístico dos itens lexicais, é possível fazer inferências sobre seu conteúdo, ambiente de ocorrência e preferência sintático-semântica. Alguns autores da LC (SARDINHA, 1999a, 2000c; STUBBS, 2001, entre outros) procuram definir os padrões lexicais de uma palavra, a fim de encontrarem no uso da língua informações que revelem a maneira como as palavras devem ser utilizadas e definidas. Na maioria das vezes, é através destes padrões que os pesquisadores têm a possibilidade de identificar ambientes lingüísticos importantes para o significado das palavras.

1.2 Padrões lexicais: o que são?

De acordo com Oliveira (2004), o padrão lexical é a forma regular ou repetida como as palavras ou estruturas ocorrem em um conjunto de textos, ou, de outro modo, os padrões lexicais são todas as palavras e estruturas regularmente associadas com uma palavra e que contribuem para o seu significado. Para a autora (2004), um padrão lexical pode ser identificado se uma combinação de palavras ocorre com relativa frequência, se é dependente de uma escolha específica de palavras e, mais, se há um claro sentido associado a ele.

Em LC (SINCLAIR, 1991; PARTINGTON, 1998; SARDINHA, 1999a), as ocorrências dos padrões se distribuem em três grandes grupos: colocação, coligação e prosódia semântica. A colocação baseia-se no item lexical em associação ao co-texto: refere-se a palavras que fazem “companhia” uma para outra, ou que geralmente co-ocorrem em uma frase ou texto. Por exemplo, a palavra ‘causar’ tende a (co)-ocorrer de maneira mais freqüente com palavras tais como ‘problemas’, ‘prejuízo’, ‘danos’, ‘morte’, ‘impacto’, como mostra o exemplo:

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

Além da co-ocorrência, os itens lexicais podem associar-se gramaticalmente. Os autores acima denominam essa associação de coligação. Mais especificamente, a coligação é a companhia gramatical mantida pelo item lexical e as posições que ele “prefere” na sentença. Um exemplo de coligação fornecido por Sardinha, (2000b:2), tomando por referência a palavra “só” num estudo de caso do português, ilustra essa situação:

02a) Eu **só** vi alguns cartazes por aí.

02b) O PFL prevê que a tensão **só** vai diminuir após a volta de FHC.

02c) Ninguém lá faz benfeitoria. **Só** fazem filhos e feitiçaria.

Em (02a), o sentido é o de restrição, exclusão, seleção; em (02b), ‘só’ assume um sentido de condição; finalmente, em (02c), o sentido veiculado é de crítica. Há outros sentidos associados à palavra ‘só’, de acordo com o posicionamento ou sentido

dos itens colocados, conforme explicitado por Sardinha (2000) em sua análise. Por exemplo, a coligação proposta pelo autor, 'pode+ser+particípio', é empregada para indicar adversidade como no exemplo:

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

Destaca-se que a interpretação dessas sentenças depende da carga semântica dos verbos que acompanham o padrão da coligação. É necessário, portanto, reparar na interligação dos padrões para que se produza significado.

Além da colocação e da coligação, os itens lexicais também podem se associar através de um uso freqüente em determinado campo semântico³: a prosódia semântica (SARDINHA, 1999a), que pode ser negativa, positiva ou neutra. A sua determinação não só depende do que a própria palavra traz em si, mas também do próprio contexto em que aparece. A título de ilustração, a palavra 'acontecer' tem uma prosódia semântica que poderá ser tanto negativa quanto neutra, dependendo do contexto em que se insere. Por exemplo, quando acompanhada de palavras como 'coisa' e 'algo' (talvez 'caso'), será claramente neutra:

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

Neste exemplo a palavra 'algo', colocada do núcleo 'aconteceu', não sobrepõe ao exemplo uma carga negativa ou positiva, estabelecendo uma semântica neutra à

³ Define-se campo semântico como o conjunto de lexemas ou unidades lingüísticas semanticamente relacionadas, tanto ao nível paradigmático quanto ao sintagmático, dentro de um dado sistema lingüístico. (LYONS, 1977a:217).

sentença. Porém, se 'acontecer' estiver acompanhada de palavras altamente negativas como 'crime' e 'acidente', o efeito é negativo:

04) Um *crime* horrível **aconteceu** no bairro noite passada.⁴

Reforçamos, aqui, a importância dos subsídios fornecidos pelos três padrões para a interpretação dos textos. Essa relação entre os padrões e as informações que podem ser percebidas a partir deles é possível porque o nosso conhecimento não está relacionado a palavras individuais, mas às suas combinações e ao 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 dessas combinações a identificação do assunto do texto. O autor parte da idéia de que o significado está disperso entre as palavras que co-ocorrem no texto e que elas compartilham características semânticas (ibid., 2001:63). Ou seja, a partir das palavras e seu ambiente, é possível descrever significados mais extensos. Os padrões lexicais da colocação, coligação e prosódia semântica, assim como a distribuição das palavras observadas em um *corpus*, podem fornecer dicas importantes para o auxílio à identificação dos tópicos dos textos, proporcionando uma base empírica para a sua interpretação (STUBBS, 2001).

Até aqui, tivemos como preocupação fazer um resgate teórico de como o processo de interpretação de textos ocorre na perspectiva de alguns autores. A partir dos apontamentos realizados neste capítulo com base nas postulações de Eco (1979), Kleiman (2001, 2002), Sardinha (1999a) e outros, pode-se dizer que as palavras mantêm relações que devem ser percebidas para que se possa extrair do texto

⁴ Exemplos de nossa autoria.

informações relevantes para sua efetiva interpretação. De acordo com autores, para que se possa interpretar um texto, determinadas palavras (denominadamente palavras-chaves) assumem um ambiente capaz de fornecer pistas importantes para a interpretação dos textos e conseqüente extração de informação dos mesmos.

Procuramos, ainda, esclarecer que os padrões lexicais da colocação, coligação e prosódia semântica poderiam influir nesse processo, fornecendo subsídios importantes para a identificação do assunto do texto, bem como, algumas das informações relativas a esse assunto.

Reiteramos aqui, que o nosso objetivo com esse trabalho é investigar qual a relação entre as palavras e verificar quais aspectos nessa relação podem auxiliar o leitor no processo de interpretação de textos, e que não nos preocupamos neste momento em apresentar uma aplicação pedagógica para o método. Porém, uma discussão formal sobre aspectos lingüísticos será fornecida no decorrer do trabalho que pode servir de base para futuros trabalhos que tenham a intenção de encontrar uma aplicação específica, seja ela, pedagógica ou para formalização de sistemas computacionais com objetivos de processos interpretativos.

No próximo capítulo, buscamos enfocar mais especificamente algumas das principais técnicas aplicadas ao processo de interpretação de textos.

2 TÉCNICAS APLICADAS À INTERPRETAÇÃO DE TEXTOS

As técnicas utilizadas como auxílio à interpretação de textos são variadas e dependem, dentre outros fatores, das estratégias próprias de cada leitor. De acordo com Kleiman (2001:50), essas estratégias podem ser de dois tipos: cognitivas ou metacognitivas. As primeiras são ações mentais inconscientes, que o leitor nem sempre é capaz de descrever. A utilização do conhecimento prévio é um exemplo delas. As estratégias cognitivas incluem, também, as habilidades lingüísticas do leitor, sem as quais, como vimos no capítulo 1, é impossível compreender um texto. Para a autora (2001:65), essas estratégias

se apóiam, basicamente, no conhecimento das regras gramaticais (regras sintáticas e semânticas de nossa gramática interna ou implícita, que não tem nada a ver com a gramática escolar) e no conhecimento de vocabulário (que subjazeria de alguma forma ao reconhecimento instantâneo das palavras do texto).

As estratégias metacognitivas, por sua vez, referem-se às operações sobre as quais o leitor tem controle no ato da interpretação, podendo explicá-la a outra pessoa:

as estratégias metacognitivas seriam aquelas operações (não regras), realizadas com algum objetivo em mente, sobre as quais temos controle consciente, no sentido de sermos capazes de dizer e explicar nossa ação. (KLEIMAN, 2001:50)

Um exemplo desse tipo de estratégia é quando o leitor percebe que não entendeu o texto e decide reler, procurar significados para palavras desconhecidas através do contexto, ou fazer um resumo.

De acordo com a autora, para auxiliar o leitor a interpretar adequadamente, é necessário desenvolver ações que atuem sobre ambas as estratégias, a fim de mostrar-lhe quais caminhos seguir para construir um sentido do texto. Sua proposta consiste, “por um lado na modelagem de estratégias metacognitivas, e, por outro, no desenvolvimento de habilidades verbais subjacentes aos automatismos das estratégias cognitivas” (KLEIMAN, 2001:50). É no interior do ato de leitura, enfrentando os desafios colocados pelo texto, que se aprende a usar as estratégias de leitura como ferramenta para desenvolver técnicas próprias e eficientes para o processo interpretativo (ibid., 2001). Utilizando estes dois tipos principais de estratégias, o leitor procuraria extrair a informação do texto.

Há diversas maneiras de levar o leitor à ativação de estratégias de compreensão. Em âmbito escolar, por exemplo, uma das técnicas mais utilizadas é a de fazer perguntas relacionadas ao texto a fim de checar a compreensão dos alunos. Nesta tentativa, o professor pode questionar os alunos sobre alguns conhecimentos prévios que podem ser necessários para a compreensão de um texto, a fim de levá-los a ativar esses conhecimentos e prepará-los para receber as novas informações durante o processo de leitura.

Ainda que, talvez, inconscientemente, todo professor sabe que os seus questionamentos devem ir além do pressuposto de que a sua interpretação do texto é a única coerente e “correta”; sendo assim, as perguntas sobre o texto não devem ser feitas de maneira a levar os alunos a identificar unicamente as informações explícitas. Isso porque o processo de interpretação, conforme discutido no capítulo 1, consiste no encontro de informações implícitas dadas não só pelas relações estruturais, mas também pelas relações semânticas estabelecidas no texto. A discussão sobre questões

meramente óbvias faz com que o aluno seja um simples reprodutor das frases do texto, o que ele pode fazer com a simples identificação de frases soltas, mas sem a compreensão completa. Muitas vezes, e, como dissemos, talvez até sem perceber, o que esse professor faz é conduzir os seus alunos a constituírem os *frames* citados por Eco e, de certo modo, retomados por Kleiman.

Uma outra técnica bastante utilizada pelos professores é a chamada “*cloze*” (ALLIENDE & CONDEMARIN, 1987). Basicamente, a técnica consiste na omissão de algumas palavras de um texto, que o aluno deverá completar com alguns possíveis itens lexicais. Para tanto, parte-se do pressuposto de que as palavras geralmente mantêm uma ambiência de ocorrência em determinados contextos. O contexto e as pistas léxico-semânticas deixadas no texto possibilitam a identificação dos vocábulos omitidos ou a inferência sobre o sentido exposto. Para que esta técnica ocorra de maneira adequada, o leitor deve ser capaz de perceber a função do contexto na ativação de um item lexical. Em tal atividade, o leitor é levado a estabelecer um determinado grau de co-relação do contexto com determinada palavra a fim de ativá-la eficientemente e estabelecer o sentido global do texto. O nível de familiaridade do leitor com o assunto abordado é imprescindível, pois desta forma o contexto em que cada palavra ocorre já está presente no seu léxico.

Decidir um título mais adequado para o texto é uma outra técnica: um título bem elaborado leva o leitor a formular hipóteses e expectativas de leitura que o auxiliam enormemente no processo de compreensão (PEREIRA, 2003; SPYRIDAKIS, 1989; KLEIMAN, 2001). Assim, ler o texto, compreendê-lo e dar-lhe um título condizente com seu conteúdo é, seguramente, uma atividade que faz o leitor refletir sobre o verdadeiro significado contido naquele texto. Da mesma forma, identificar o título e através dele

elaborar previsões e expectativas de interpretação pode constituir-se como uma atividade importante para a ativação de conhecimento prévio do aluno e a formação de hipóteses de leitura.

Fazer marcações, tais como sublinhar **sentenças ou itens lexicais** considerados importantes e fazer anotações de margem, destaca-se como mais uma técnica. Conforme Pereira (2003), essa técnica é conduzida, na maioria das vezes, intuitivamente. O professor, ao propor esse tipo de atividade, parte do pressuposto de que o reconhecimento das idéias mais importantes do texto já está implícito no ato da leitura. Para a autora,

sublinhar, marcar um texto, fazer esquemas e sínteses são realizados, normalmente de forma bastante intuitiva, sem muita reflexão e quem domina a leitura simplesmente encontra como fazer (no entanto, o leitor inexperiente apresenta grandes dificuldades nessa tarefa), por isso, freqüentemente, são tomadas como atividades que não precisam ser ensinadas (principalmente o grifar), as pessoas desenvolvem sozinhas tal capacidade e, quando realizam a tarefa a contento, demonstram que sabem ler. (PEREIRA, 2003:73).

Para realizar esta tarefa, o aluno se vê forçado a localizar as idéias que nortearam o desenvolvimento do texto. Encontrar essas idéias depende, acima de tudo, da focalização do objetivo da leitura, e não da mera atividade de “leia e sublinhe as idéias principais”.

Uma técnica que também parece produzir bons resultados é a de tomar notas sobre o que o leitor “acha importante” no texto, de acordo com os objetivos da leitura. Neste caso, ele é instigado a reproduzir com as suas próprias palavras o que teria entendido da passagem em questão, o que indica que ele deve apresentar um bom nível de construção do conhecimento do texto. No entanto, este tipo de atividade pode

levar à mera reprodução fiel de passagens contidas no material escrito, o que não indica que ele reformulou seu conhecimento. Por outro lado, mostrar ao leitor como inferir e levantar hipóteses pode ser um processo construtivo na compreensão textual. Se for bem orientada, esta pode ser uma técnica poderosa no auxílio ao processo interpretativo.

Entre os vários objetivos da leitura, um deles é a procura por uma informação concreta. Nesta tarefa, incluem-se necessidades específicas que motivam o leitor a encontrar no texto respostas às suas buscas e retirar dele a informação pretendida. Por isso, o professor ajuda em grande parte a atividade interpretativa quando estabelece um objetivo claro anterior ao processo efetivo de leitura, propiciando à mente do leitor a seleção de alguns aspectos mais importantes do texto e a desconsideração de outros.

A seleção de índices importantes para a compreensão é uma das características do leitor proficiente. Para fazer essa seleção, ele precisa elaborar os seus próprios caminhos cognitivos com o auxílio de técnicas que estabeleçam objetivos claros durante a leitura, favorecendo a procura por informações específicas. Mostrar para o leitor o que deve ser extraído do material que se propõe a ler leva-o a construir suas estratégias de compreensão mais facilmente.

Encontrar as principais idéias de um texto não é tarefa fácil, pois é uma atividade que demanda orientação e organização. Um exemplo de mediação seria a análise, na fase de pré-leitura, dos itens lexicais presentes no material (PEREIRA, 2003). Esta análise possibilita aos leitores a formação de um léxico específico, o qual será mais facilmente ativado durante a leitura, auxiliando o processo compreensivo.

Para direcionamentos da análise de itens lexicais, uma técnica que pode ser utilizada é a extração de palavras-chaves, ou, mais exatamente, daquelas que refletem

o principal conteúdo do texto (SARDINHA, 1999c; SINCLAIR, 1991; PARTINGTON, 1998). Uma área de investigação que utiliza o reconhecimento de palavras-chaves e, a partir delas, realiza a identificação das informações contidas em um texto, é a chamada Extração de Informação (EI).

2.1 Extração de Informação

De acordo com Riloff (1999:435), a extração de informação (EI) é “uma forma de processamento da linguagem natural na qual certos tipos pré-definidos de informação devem ser reconhecidas e extraídas de um texto”. Isto significa dizer que um sistema de extração de informação analisa textos e produz para eles uma representação da informação considerada relevante para uma determinada aplicação.

Para que qualquer informação seja extraída de textos, o sistema precisa ter definido previamente as tarefas a serem realizadas, o domínio e a finalidade. As tarefas podem variar conforme o texto: pode ser a simples identificação de nomes, locais, ou acontecimentos e, ainda, a localização de extratos importantes sobre um determinado assunto, o que inclui a interpretação do texto.

No caso de sistemas de EI com objetivos de interpretação, todos os eventos presentes no texto são retirados através de palavras-chaves, que, em última instância, constituem o núcleo dos padrões de extração: os núcleos lexicais. Por exemplo, em um conjunto de artigos jornalísticos sobre “questões ambientais relacionadas à proteção da vida selvagem”, é necessário retirar daí informações relativas a “quem é a favor ou contra isso?”, “a situação A está piorando ou melhorando?”, “quem é responsável pelos projetos?”, etc. A resposta a essas perguntas é dada pela ativação de sentenças e

padrões por intermédio de palavras-chaves que são previamente analisadas e descritas para o sistema de acordo com seu comportamento lingüístico, por meio da análise de *corpora*.

Para que o processo de EI seja eficaz, é essencial um trabalho de explicitação de elementos lingüísticos para a descrição dos padrões usados no sistema. Faz-se necessária a determinação das relações lingüísticas que precisam ser capturadas e de como elas estão expressas em textos de língua natural. Num conjunto de textos do domínio a ser analisado, os padrões são descritos a partir de um núcleo lexical, que revela, por seu ambiente lingüístico, quais informações devem ser extraídas.

De forma geral, os sistemas de EI buscam utilizar diversas abordagens lingüísticas para tratar as informações, embora na grande maioria das vezes, tomam por base a análise sintática, tal como mostrado no exemplo abaixo ilustrado:

01) O parlamento foi **atacado** pelos guerrilheiros.⁵

Considerando que a sentença integra um conjunto de textos no domínio do terrorismo, o objetivo aqui seria extrair o alvo de um ataque, tendo como palavra central o verbo 'atacar' (que funciona, neste caso, como *trigger*, ou seja, palavra que ativa o padrão). Um dos padrões lingüísticos pré-definidos para o verbo 'atacar', do ponto de vista sintático, que auxiliaria nesta determinação seria o seguinte:

<sujeito> verbo na voz passiva

⁵ Riloff, 1999. Tradução minha.

Se a sentença contém o verbo 'atacar' e se apresenta na voz passiva, este padrão é automaticamente ativado pelo sistema e associado com a sentença. Em seguida, o sujeito é extraído como o alvo do ataque:

O parlamento <sujeito> alvo

Porém, há vários padrões associados à mesma palavra, pois as relações por ela mantidas não são sempre iguais. Para o processamento, o verbo 'atacar' se torna central e todos os padrões lingüísticos relacionados a ele são descritos previamente. Um segundo exemplo demonstra alguns dos padrões que possivelmente poderiam ser descritos:

02) Os guerrilheiros **atacaram** o parlamento.

<sujeito> verbo na voz ativa (retirada do sujeito como o culpado do ataque = guerrilheiros)

verbo na voz ativa <objeto direto> (o objeto direto como alvo = parlamento)

A mera análise sintática não é suficiente, porque a língua é dinâmica e comporta-se diferentemente de acordo com diferentes contextos. Por exemplo, se a sentença for

03) O parlamento foi **atacado** e a organização guerrilheira diz ser a responsável.

não há uma referência sintática direta aos criminosos. Somente por meio de inferências e do entendimento da sentença (às vezes, até de parágrafos) é possível ligar o papel de criminoso à organização guerrilheira. De acordo com Riloff (1999),

em muitos casos, padrões locais não são suficientes, e a sentença inteira deve ser entendida. E, às vezes, inferências devem ser geradas pela observação das sentenças ou até mesmo, parágrafos. Inferências podem ser essenciais para determinar a relevância de determinada informação.⁶

O estabelecimento de inferências depende da ativação de conceitos relacionados às expressões lingüísticas específicas no domínio estabelecido.

Conforme Cowie & Wilks (2000), a definição de padrões envolve, ainda, a seleção de elementos de informação necessários, bem como a definição de seus relacionamentos sintáticos, a descrição de uma estrutura lingüística e instruções detalhadas na determinação do conteúdo dos elementos que preenchem as lacunas sintáticas.

Uma vez que a informação desejada é geralmente restrita a um conjunto de relações e eventos em um certo domínio, tipicamente uma análise gramatical é considerada suficiente pela maioria dos sistemas construídos. Restrições semânticas são estabelecidas para os constituintes que completam as lacunas (*slots*) dos padrões, porém, na maioria das vezes, isso é feito de maneira superficial. Na sentença-exemplo dada em (01), a restrição semântica que se institui é que o sujeito deve ser objeto físico. Assim, um padrão apresenta uma preferência forte para somente um determinado subconjunto das categorias que poderiam preencher as suas lacunas.

⁶ Ellen Riloff, 1999: 440. Tradução minha.

Nem tudo o que, à primeira vista, possa ser tomado como um padrão o será na realidade. Critérios para o seu estabelecimento são determinados pela literatura. Segundo Yangarber & Grishman (2000), por exemplo, o primeiro é que tais padrões devem ser prováveis de co-ocorrerem dentro de um dado documento; outro é que alguns deles devem ser similares a outros padrões, e, finalmente, os padrões que envolvem os mesmos argumentos devem ser focos de atenção, uma vez que os fatos podem ocorrer mais de uma vez no texto.

Cardie (1993) é mais específico com relação aos critérios. Para ele, três pontos seriam os fundamentais a serem levados em consideração:

- i) Características de definição da palavra: conhecimento semântico e sintático das palavras; mais especificamente, as características da palavra devem ser examinadas, rotulando-a com partes do discurso, sentido geral e específico, e morfologia da palavra.
- ii) Característica de contexto local: após o primeiro passo, devem ser analisadas as duas palavras anteriores e as duas seguintes à palavra em questão, utilizando-se também o processo de identificação sintática e semântica das palavras.
- iii) Características de contexto global: deve-se codificar a informação, incluindo conhecimento sintático e semântico para os constituintes da sentença num geral.

O reconhecimento de constituintes e as características da frase, assim como a identificação da classe das palavras, auxiliam no processo de construção dos padrões gerais.

A descrição bem fundamentada (principalmente em bases empíricas) dos padrões é que vai determinar a eficácia do sistema. Assim, a observação e a análise de dados empíricos são essenciais na definição de bons padrões de EI. A lingüística de *corpus*, por exemplo, pode oferecer bons subsídios para a análise dos padrões de uma palavra, pois trabalha com dados concretos e textos em língua natural, além de fornecer propostas teóricas bem fundamentadas para a análise lingüística dos textos.

Neste capítulo, focalizamos uma discussão sobre as técnicas utilizadas para auxiliar o leitor na interpretação de textos. Baseados em alguns autores que trabalham especificamente com interpretação de textos para ensino, como p. ex., Kleiman (2001; 2002) e Pereira (2003), nota-se que é um processo complexo, que exige atenção dos leitores para determinados aspectos lingüístico textuais.

Em especial, procuramos demonstrar o funcionamento da EI (RILLOFF, 1999, e outros), conceito que se torna essencial para este trabalho devido à preocupação em esclarecer relações lingüísticas que possibilitem a identificação das idéias presentes em um texto.

No capítulo a seguir, explanam-se os critérios metodológicos que foram necessários para a análise aqui proposta.

3 ASPECTOS METODOLÓGICOS

A análise proposta neste trabalho tem por objetivo apontar alguns aspectos lingüísticos que contribuem para o processo de interpretação de textos. Para a sua realização, alguns critérios metodológicos se fizeram necessários e serão descritos neste capítulo da pesquisa. Consideramos, em especial, três elementos:

- (i) A seleção do *corpus* de pesquisa
- (ii) A seleção do *corpus* de referência
- (iii) A seleção das palavras-chave.

3.1 A seleção do *corpus* de pesquisa.

Denomina-se *corpus* um “conjunto de textos designados para pesquisa lingüística” (STUBBS, 2001:25). É por meio da observação de um *corpus* que o pesquisador pode checar suas intuições com relação à língua, e verificar que informações lingüísticas são relevantes de acordo com seu objetivo (PARTINGTON, 1998).

O motivo que nos leva a fazer uma análise baseada em *corpus* é que os significados ficam evidentes no contexto lingüístico investigado em um determinado conjunto de textos representativos da língua, o que seria um ponto importante a ser considerado na interpretação de textos.

Para a escolha do *corpus* desta pesquisa, os seguintes critérios foram observados:

- i) Os textos deveriam ser informativos e dirigidos a um público não especializado;

- ii) A linguagem usada nos textos deveria ser de fácil acesso a um número variado de pessoas, isto é, estar facilmente disponível a qualquer pessoa que se interessasse pelo assunto;
- iii) Todos os textos deveriam ser assinados por um profissional da área.

Aplicados esses critérios, 61 textos foram selecionados, totalizando 49.088 palavras. Todos os textos foram extraídos da internet (revistas eletrônicas e *sites* especializados) e pertencem à área da saúde, mais especificamente, da gastroenterologia. Os textos possuem um perfil informativo, de interesse geral para toda pessoa que procura informações sobre sintomas, causas, tratamentos de problemas relacionados ao sistema digestivo. Ou seja, o assunto principal dos textos está relacionado à descrição de problemas gástricos.

Uma vez selecionados, os textos foram submetidos a uma pré-análise, por meio do programa *Wordsmith Tools*⁷, em que as palavras são relacionadas conforme a frequência de ocorrência. De acordo com Sinclair (1991), a observação da frequência das palavras e o uso de listas de palavras e conjuntos de suas distribuições nos dão uma primeira idéia de como um texto está organizado, fornecendo, assim, uma base empírica para a interpretação.

A fim de localizar, neste grupo de textos, as palavras núcleo (palavras-chave) que foram utilizadas neste trabalho, fez-se necessária, antes, a seleção de um *corpus* de referência.

⁷ Ferramenta disponível em versão demonstrativa no site: www.lexically.net/wordsmith

3.2 O *corpus* de referência

O *corpus* de referência é o conjunto de textos que servem de base para a comparação com o *corpus* de pesquisa, a fim de retirar deste as palavras-chaves (SARDINHA, 1996). Uma das características do *corpus* de referência é a de que ele deve estar no mesmo domínio do *corpus* de pesquisa. Assim, para a determinação desse *corpus*, especialistas da gastroenterologia foram consultados.

Outro critério para sua seleção é que o *corpus* de referência deve ser maior que o de pesquisa. Neste sentido, este conjunto de textos compreende um total de 97.372 palavras retirados de um livro da área de gastroenterologia (Dani, 2001) o qual, de acordo com o especialista consultado, constitui uma boa base de pesquisa.

O mesmo processo de pré-análise utilizado para o *corpus* de pesquisa é realizado com o *corpus* de referência. Faz-se, assim, uma lista de frequência das palavras deste *corpus* para a posterior comparação e retirada das palavras-chaves.

3.3 O papel das palavras-chaves na interpretação de textos.

Palavras-chaves, por definição, são palavras que possuem uma frequência estatisticamente superior num *corpus* em comparação a um *corpus* de referência, e que em conjunto refletem a idéia principal do conteúdo do texto (SARDINHA, 1999a). Para o autor, uma das várias possibilidades de estudos lingüísticos baseados em palavras-chaves seria a de identificar, por meio delas, os possíveis conteúdos discutidos em um texto.

O que Sardinha denomina palavras-chaves, para Cavalcanti (1989) são itens lexicais chaves. A autora parte do pressuposto de que problemas de interpretação estão relacionados a sinais-chave: os itens lexicais chaves. Para esta autora,

os itens lexicais chaves são um elo de ligação entre leitor e texto e entre as suas características; a partir deles e da sua relação com outros itens lexicais do texto, molduras, esquemas e valores são ativados pelo leitor. O conceito de um elemento chave no texto depende imediatamente dos outros itens a ele associados, fazendo com que a integração das palavras estabeleça um sentido para o texto. (CAVALCANTI, 1989:81).

Cavalcanti acrescenta que os itens lexicais chaves são expressões cujo significado baseia-se no contexto no qual eles são usados, ou seja, são caracterizados na relação com outros itens identificados no texto. O argumento da autora, o qual adotamos neste trabalho, é o de que

itens lexicais chaves centralizam informação que tem relações diretas no texto e relações pragmáticas indiretas na interação leitor-texto. Eles servem como espinha dorsal para o estabelecimento de conteúdo proposicional e força ilocucionária, ativam estruturas de conhecimento (esquemas ou *frames*) e os sistemas de valores do leitor; e subjazem à construção do pressuposto em relação ao todo ou parte do texto. (CAVALCANTI, 1989:75)

O que queremos dizer com isso é que a observação de itens lexicais chaves é essencial no processo interpretativo porque ativam os esquemas (*frames*) propostos por Eco e, de certa forma, retomados por Kleiman, e ainda centralizam a informação objetivada pelos processos de extração de informação.

Associados aos itens lexicais chaves, aparecem outros itens (denominados por Cavalcanti associativos, contextualizantes e iterativos), os quais são responsáveis pela

indicação do conteúdo semântico, da coesão textual, de como as orações se relacionam entre si e, ainda, servem como ponto de apoio para o leitor no estabelecimento de força ilocucionária e conteúdo proposicional do texto.

Para que um item seja considerado chave, na visão de Cavalcanti (1989:82), eles devem apresentar duas propriedades básicas: a propensão à saliência e a direcionalidade de âmbito. A primeira é a propriedade de um item lexical de ser localizado e repetido nas partes importantes do discurso, de servir de referente para manter a coesão gramatical, de ser associado a outros itens lexicais chaves (colocações associativas explicadas por Scott, 2004⁸), e de ser incluído como membro de relações clausais. A segunda propriedade dos itens lexicais chaves é a de direcionar a ativação de esquemas e de sistemas de valores do leitor. É ela que permite a delimitação do tópico e a determinação das intenções da comunicação presentes no texto.

Sardinha (1999c) destaca que as palavras-chaves permitem descobrir quais as possíveis informações presentes em um determinado texto. Para este autor, as colocações associadas às coligações e à prosódia semântica podem fornecer uma melhor imagem dos significados específicos associados às palavras-chaves em contexto. Por isso, destacamos a importância de analisá-las no processo interpretativo.

3.3.1 A seleção das palavras-chaves

Para a seleção das palavras-chaves, levamos em consideração ainda a sua classe gramatical. Via de regra, são as palavras de classe aberta as que tendem a

⁸ Os *Associates* (SCOTT, 2004), ou itens associativos (CAVALCANTI, 1989), serão retomados no tópico 4.1 do presente trabalho.

revelar a idéia central do texto. Por este motivo, opta-se pela seleção, observação e análise destes itens. Isso não quer dizer que as palavras de classe fechada sejam menos importantes; porém, por terem um caráter mais restrito, a contribuição que fornecem para a interpretação é bem menos significativa se comparadas às de classe aberta.

Conforme exposto na seção 3.1, um primeiro critério utilizado para a seleção das palavras-chaves foi a freqüência de ocorrência, ou seja, uma das características das palavras-chaves é, de fato, a de ser repetida várias vezes no decorrer dos textos. Mas a freqüência, por si só, não se sustenta como critério único de seleção das palavras. Por este motivo, para o presente estudo, as características das palavras obtidas na amostra foram analisadas de acordo com a proposta de Cavalcanti (1989) sobre itens lexicais chaves, discutida na seção anterior deste trabalho (veja-se 3.3).

Em suma, todas as palavras do *corpus* de estudo foram submetidas a um tipo de análise. As primeiras análises, como, por exemplo, o cálculo de freqüência e as comparações dos *corpora*, foram realizadas com o auxílio do programa *Wordsmith Tools*. Também por meio de ferramentas do programa, as palavras-chaves foram identificadas e posteriormente submetidas a uma nova filtragem, com o intuito de selecionar algumas delas para o presente estudo. A descrição dos critérios utilizados nesta filtragem, bem como as análises dos dados obtidos, serão descritas no capítulo a seguir.

4 UMA ANÁLISE APLICADA À INTERPRETAÇÃO DE TEXTOS

4.1 Sobre a filtragem das palavras-chaves

O *software* de auxílio utilizado nesta pesquisa, *Wordsmith Tools*, retira do *corpus* de pesquisa todas as palavras-chaves por meio de um cálculo estatístico em comparação com o *corpus* referência. No caso deste trabalho, o resultado mostrado pelo programa foi um total de 94 palavras-chaves (positivas)⁹. No entanto, seria impossível para um pesquisador analisar todo esse conjunto de palavras. Como Sardinha afirma (1999c:4),

a lista de palavras-chave produzida pelo programa normalmente tem mais palavras-chave do que é possível para um pesquisador analisar. Como resultado, a maioria dos pesquisadores seleciona uma porção do total de palavras-chave para analisar.

Assim, com o intuito de selecionar a porção de palavras para a nossa análise, a lista de palavras-chaves obtidas se submete a um novo processamento, no qual todos os textos são analisados separadamente pelo programa, e, como resultado, obtém-se uma lista das chamadas “palavras superchaves”.¹⁰

As palavras superchaves são aquelas que aparecem como chave em mais de um ou em vários dos textos da pesquisa (SCOTT, 2004).¹¹ A lista de palavras é mostrada pelo programa de acordo com o número de textos em que cada palavra é chave, a respectiva porcentagem e as demais palavras-chaves associadas à palavra

⁹ Palavras-chaves positivas são aquelas que apresentam frequência estatisticamente maior no *corpus* de estudo comparado com o *corpus* de referência. (SCOTT, 2004)

¹⁰ Do inglês, *key keywords*.

¹¹ Tradução minha.

superchave. A tabela abaixo ilustra as primeiras (isto é, as que aparecem em um maior número de textos) das 75 palavras superchaves mostradas pelo programa:

Tabela I: Lista de palavras superchave mostradas pelo programa *Wordsmith Tools*¹²

Palavra-chave	Número de textos em que a palavra aparece como chave	Frequência de textos em que aparece como palavra-chave (%)	Associadas
INTESTINO	8	15	Você, digestão
PYLORI	7	13	
VOCÊ	7	13	Intestino, digestão
DIGESTÃO	5	9	Intestino, você
BACTERIA	4	7	Pylori
CAUSA	4	7	
ALIMENTOS	3	5	
BOCA	3	5	Você
DUODENO	3	5	
ESTOMACAL	3	5	
ALIMENTO	2	3	Você
ANTIBIÓTICO	2	3	
EXAME	2	3	

De acordo com Scott (2004), quanto mais vezes uma palavra aparecer como chave, mais relevante ela é, e mais ela indicaria o assunto tratado nos textos. Com base nessa constatação, optou-se por selecionar palavras que apareciam como chave no mínimo em quatro dos textos analisados, ou seja, em 7% dos textos. Tendo em vista os critérios estabelecidos acima, a seleção a que procedemos resultou em cinco palavras a serem analisadas: intestino, *Pylori* (*H pylori* ou *Helicobacter pylori*), digestão, bactéria e causa. A palavra ‘você’ é descartada por se tratar de uma palavra de classe fechada.

Observando-se a última coluna da tabela I acima, vemos que algumas palavras superchaves apresentam palavras associadas. Tais palavras são também palavras-

¹² A tabela completa pode ser visualizada no anexo B do presente trabalho.

chaves que geralmente se ligam à palavra superchave nos mesmos textos (SCOTT, 2004). Poderíamos dizer que estas palavras participam do significado das palavras superchaves, ou seja, revelam o mesmo ambiente lingüístico de ocorrência. Portanto, as palavras 'intestino' e 'digestão' fazem parte do mesmo contexto, assim como as palavras '*pylori*' e 'bactéria'. Diante desta constatação, optamos por realizar a análise dessas palavras de forma concomitante. Assim, poderão ser mais freqüentemente encontradas na seqüência deste trabalho, três palavras analisadas: 'intestino', 'causa' e '*helicobacter pylori*'. Isso não significa que as palavras 'bactéria' e 'digestão' tenham sido desconsideradas. Ao contrário, a análise a que nos propomos realizar não se limita unicamente às palavras superchaves destacadas, mas leva em consideração o seu contexto lingüístico de ocorrência. Por isso, tais palavras aparecerão como correlatas na análise para auxiliar na interpretação dos resultados.

A fim de que o contexto lingüístico fosse melhor observado, retiraram-se, por meio da ferramenta "concordanciador" do programa, todas as ocorrências das palavras superchaves selecionadas do *corpus* de pesquisa em uma janela de 12:12, ou seja, 12 colocados à esquerda e 12 à direita de cada palavra (Vide Anexo A).

Conforme vimos anteriormente, para que as palavras-chaves revelem conteúdo e forneçam pistas para sua interpretação, é necessária a avaliação dos outros itens a elas associados. Assim, tendo por base os padrões lexicais da colocação, da coligação e da prosódia semântica, as ocorrências em contexto de cada uma das palavras selecionadas foram analisadas. Os resultados podem ser melhor observados nas próximas seções deste trabalho organizados nas tabelas II, III e IV, a partir das quais uma descrição dos resultados será realizada com mais profundidade.

4.2 Palavras-chaves e padrões lexicais: revelações para a interpretação de textos.

Reforçamos, aqui, que uma das nossas principais preocupações neste estudo é verificar de que maneira as relações léxico-semânticas mantidas pelas palavras-chaves fornecem informações importantes para a interpretação de textos. As análises realizadas sobre um pequeno, mas representativo, grupo de palavras-chaves apontam que elas revelam um ambiente lingüístico bem marcado, construindo em torno de si situações semânticas altamente importantes para o processo interpretativo.

Por meio da análise de cada uma das palavras-chaves selecionadas, destacaram-se, dentre outros aspectos, as diferentes situações semânticas assumidas pelos itens lexicais individualmente. Em todas, constataram-se indícios reveladores do assunto tratado nos textos, bem como algumas das possíveis interpretações para os mesmos. Com base nas porções textuais em que cada palavra-chave ocorria, procuramos identificar como a relação entre as palavras acontecia e que tipo de informação se podia extrair dessas relações.

Conforme vimos anteriormente, várias palavras-chaves foram rastreadas. Embora coubesse o registro do que se pôde verificar com relação a elas em suas diferentes manifestações, optamos por promover uma redução do material coletado devido à enorme quantidade de informação gerada. A seguir, serão apresentados os resultados das análises feitas sobre as palavras escolhidas.

4.3 As relações contextuais da palavra ‘intestino’

Uma das palavras selecionadas para a análise foi ‘intestino’. Para essa e para todas as palavras discutidas a seguir, o procedimento adotado teve como base descobrir não só o número de ocorrências, mas também os contextos lingüísticos em que aparece.

No caso da palavra ‘intestino’, a tabela II abaixo fornece um demonstrativo com base na observação dos padrões lexicais estabelecidos pela palavra. Do *corpus* analisado, foram contabilizadas 135 ocorrências de ‘intestino’, assim distribuídas:

Tabela II: Padrões lexicais mantidos pela palavra ‘intestino’.¹³

OCORRÊNCIAS DA PALAVRA INTESTINO		
COLOCAÇÕES	COLIGAÇÕES	PROSÓDIA SEMÂNTICA PREDOMINANTE
Delgado (53) Grosso (32) Porção (14) Duodeno (11) Parte (10) Cólon (9) Alimentos (7) Digestão (6) Estômago (6)	Substantivos de campo semântico semelhante à palavra (estômago, duodeno) Adjetivos que levam à formação de um substantivo composto (intestino grosso, intestino delgado) Verbos que indicam direção (ex. passar, continuar)	Negativa ou neutra

Tanto pela colocação quanto pela coligação, a palavra ‘intestino’ se relaciona com diversas palavras, permitindo extrair diferentes resultados interpretativos por meio da observação de cada um desses padrões em particular. A prosódia semântica nos remete a uma expectativa negativa, ou, em alguns casos, neutra, dependendo da

¹³ O número ao lado das colocações indica a quantidade de vezes em que o colocado aparece no mesmo ambiente da palavra-chave em análise.

situação, devido à carga semântica estabelecida pelas palavras ao seu redor, como veremos com mais profundidade deste ponto em diante do trabalho.

Quando a palavra-chave ‘intestino’ aparece com colocados tais como ‘delgado’ e ‘grosso’, por exemplo, o efeito é nitidamente de restrição. O objetivo nessas ocorrências está claramente centrado no órgão ‘intestino’, que tem uma de suas partes citadas de maneira mais restrita, como exemplificado a seguir¹⁴:

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

O alvo de atenção do leitor, neste exemplo, é apenas uma parte do intestino. Este mesmo tipo de restrição também acontece quando ‘intestino’ apresenta colocação com palavras como ‘porção’ e ‘parte’. Mas, em alguns casos, o alvo pode ser ampliado, focando o sistema digestivo como um todo e não apenas o órgão citado. Essa situação é ilustrada no exemplo abaixo:

24) [...] permitindo o exame detalhado do revestimento interno da *porção* superior do trato gastrointestinal, isto é, o esôfago - estrutura que liga a boca ao estômago-, estômago e duodeno - primeira *porção* do **intestino delgado** -, usando um aparelho, o endoscópio [...]

Em (24), a palavra ‘porção’ aparece como elemento de restrição. Mas, ao mesmo tempo, a combinação com o nome de outros órgãos (por exemplo, estômago e duodeno), de certa forma, faz com que o leitor amplie seu foco de atenção para um sentido mais geral do sistema digestivo.

¹⁴ Os números usados nos exemplos são respectivos ao número da ocorrência no *corpus*, conforme anexo A.

Também é possível que as colocações com palavras que denotam outros órgãos provoquem um efeito diferente. Neste caso, na maioria das vezes, há uma forte tendência para a identificação da localização de alguma patologia. Vejam-se os exemplos:

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

31) [...] Quando se fala em *úlcera*, 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 [...]

Em ambos os exemplos, a palavra ‘intestino’ tem um papel semântico de localizador das patologias: diverticulite em (86) e úlcera em (31). Se for verdade que as palavras colocadas com as palavras-chaves é que conduzem ao significado do texto através de fios semânticos liderados pelo núcleo (CAVALCANTI, 1989), então podemos afirmar que a palavra ‘intestino’ propiciaria ao leitor ativar os esquemas relacionados a palavras que se referem a outros órgãos do sistema digestivo (duodeno, estômago) ou partes dele (cólon), estabelecendo uma ligação semântica com esses colocados. A partir dessas ligações e da ocorrência de uma palavra que denota uma patologia, o leitor ativaria uma possível interpretação relacionada à patologia, reconhecendo o local de ocorrência da doença.

Em nossas análises, outras colocações também puderam ser destacadas em relação à palavra ‘intestino’. Quando em co-ocorrência com a palavra associada ‘digestão’, há um prenúncio de um processo digestivo que está em andamento ou concluído. É o que mostra o exemplo a seguir:

54) [...] A digestão química é *completada no intestino* delgado e os constituintes da refeição são absorvidos no sangue [...]

Os colocados responsáveis pela interpretação de “processo digestivo”, além do próprio item lexical ‘digestão’, são palavras que levam o leitor a ativar esquemas de representação semântica relacionados a processo, como, por exemplo, ‘refeição’ e ‘completada em’.

Vale observar que não é só a partir das colocações com palavras da classe dos substantivos que se chega a alguma informação relevante no texto. Uma coligação de ‘intestino’ com verbos é igualmente importante. Locuções verbais que contêm traços semânticos indicativos de direção ou passagem (por exemplo, ‘passar por’, ‘ir para’), sugerem, em grande parte das ocorrências, que a palavra ‘intestino’ funciona como um ponto de passagem ou um receptor de processamentos gastro-intestinais.

O que falamos até aqui teve o objetivo de mostrar que, dependendo de como e com o que ‘intestino’ aparece no texto, diferentes tipos de informação a respeito das palavras podem ser extraídas. Análises mais acentuadas nos permitiram detectar pelo menos cinco tipos de interpretação diferentes para as porções de textos em que a palavra ‘intestino’ aparece.

4.3.1 ‘Intestino’ na interpretação de um órgão receptor.

Uma das primeiras constatações que pudemos fazer foi que a referência a ‘intestino’ pode conduzir a uma interpretação de locativo-receptor do órgão. Tomem-se por base os exemplos (8) e (22) abaixo:

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

Nesta ocorrência apresentando 'intestino' como núcleo, a palavra 'empurrar' tem um traço semântico que indica uma direção. A expressão 'em direção a' reforça essa idéia. O mesmo acontece em (22), com a presença da locução verbal 'sair para':

22) [...] 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 [...]

Neste exemplo, é o traço semântico da locução verbal 'sair para' que estabelece a interpretação de direcionamento. Tanto em (8) quanto em (22), as palavras que co-ocorrem com 'intestino' exigem entre seus colocados uma palavra que preencha a informação semântica: localização (empurrar para onde?, sair para onde?), papel esse assumido pela palavra 'intestino'.

A interpretação de locativo-receptor é obtida ainda pela trajetória dos alimentos no sistema digestivo, como mostra o exemplo (27):

27) [...] A *viagem continua até* o íleo, a última porção do **intestino** delgado, *chegando ao intestino* grosso [...]

A palavra 'viagem' e as locuções verbais 'continuar até' e 'chegar em', também por seus traços semânticos que indicam movimento, mantêm uma relação forte que institui na sentença a noção de direcionamento, com 'intestino' aparecendo ainda como

um locativo-receptor. Neste contexto interpretativo da palavra 'intestino', há um forte indício de que o assunto tratado diz respeito à descrição do processo digestivo.

4.3.2 Intestino' entendido como ponto de passagem.

Tanto quanto para o caso anterior, 'intestino' pode ainda assumir a função de órgão de passagem para objetos gastro-intestinais, como mostrado no exemplo abaixo:

47) [...] das proteínas em forma de aminoácidos e das gorduras como ácidos graxos. *Depois do intestino* grosso, as fezes *passam pelo* reto para serem eliminadas. [...]

Similarmente ao caso anterior, o que se percebe aqui é que a palavra 'intestino' vai preencher a informação de localização exigida pelas expressões 'passar por' e 'depois de' (passar por onde?, depois do quê?). O verbo indica a interpretação de movimento na oração e a expressão 'depois de' indica a idéia de continuação do processo, o que estabelece para intestino a função de ponto de passagem e não de receptor. O exemplo a seguir mostra também o sentido de passagem:

50) [...] a digestão das gorduras. O alimento digerido *passa*, então, *pelos* seis metros de **intestino** delgado remanescentes; [...]

Também neste caso, a locução verbal 'passar por', que contém um sentido de deslocamento, confirma para a palavra 'intestino' uma interpretação de ponto de

passagem. Esse colocado estabelece, ainda, na sentença o sentido de que há um caminho a ser percorrido pelos alimentos no processo digestivo.

A exemplo da primeira categoria apontada, pode-se extrair dessas ocorrências a informação de que o assunto desenvolvido nestas porções de texto é a explicação do processo digestivo como um todo. Essa interpretação é confirmada pela associação mantida entre as palavras-chaves 'intestino' e 'digestão'.

De acordo com Cavalcanti (1989), uma das características das palavras-chaves é a de indicar o conteúdo tratado no texto em que aparece por meio dos fios semânticos que mantém. Nas ocorrências de 'intestino' analisadas até aqui, a indicação de um conteúdo relacionado ao processo digestivo pode ser fortemente percebida, reforçada pela co-ocorrência mantida com a palavra associada 'digestão'.

4.3.3 'Intestino' no sentido de localizador

Além da situação semântica de locativo-receptor e de ponto de passagem, há ocorrências em que 'intestino' aparece em co-relação com palavras que denotam outros órgãos do sistema digestivo, mais especificamente com 'duodeno'. Neste tipo de ocorrência, a palavra 'intestino' tende a assumir uma função de indicador da localização do duodeno, o que pode ser observado no exemplo (38):

38) [...] A *endoscopia* é um procedimento que permite o *exame* do revestimento interno do *esôfago*, *estômago* e *duodeno* (1ª *porção* do **intestino** fino). O endoscópio é um aparelho flexível, de diâmetro menor que 1 cm [...]

A parcela de texto explicativa em que ‘intestino’ aparece (“primeira porção do intestino grosso”) está totalmente ligada à palavra ‘duodeno’ e tem a função de estabelecer, para o leitor, a posição do duodeno no sistema digestivo. Pode-se dizer que ‘intestino’, nesta explicação, ocorre para completar os traços semânticos de ‘duodeno’.

Similarmente, no exemplo (83) abaixo, a informação ‘porção inicial do intestino’ aparece também como uma parcela explicativa para indicar a localização do ‘duodeno’ no sistema digestivo:

83) [...] digestivo superior, ou seja, a boca, o esôfago, o estômago e a *porção inicial do intestino*, o *duodeno*. Nesse *exame*, o médico pode ver [...]

Chama a atenção, em muitos desses casos, o fato de haver um tipo de ocorrência direta não só com a palavra ‘duodeno’, mas também com outros órgãos do sistema digestivo (estômago, esôfago), o que aponta para o posicionamento dos órgãos em exames. Veja-se que nas duas orações há a aparição da própria palavra ‘exame’ estabelecendo essa interpretação.

Para Kato (1999), o acionamento de um esquema (por exemplo, ‘intestino’) pode levar ao acionamento sucessivo de seus subesquemas ou esquemas superordenados (por exemplo, ‘estômago’, ‘duodeno’, ‘sistema digestivo’, ‘exames’), o que levará o leitor a fazer predições importantes para uma leitura significativa. À guisa de ilustração, retomemos a ocorrência (38), em que, a partir da observação da palavra ‘intestino’ e dos seus colocados denotativos de outros órgãos e exames, é possível prever que o texto está falando de assuntos relacionados a procedimentos diagnósticos para detectar patologia e local.

4.3.4 A palavra 'intestino' em relação de causa e efeito.

Além das situações semânticas acima consideradas, 'intestino' também pode estar desempenhando um papel de receptor numa relação de causa e efeito.

Consideremos o exemplo abaixo:

64) [...] náuseas e vômitos, as mais freqüentes são as que *causam* "irritação" do estômago ou **intestino**, como as *infecções* alimentares, *gastroenterites* agudas e *medicamentos* [...]

A aparição das palavras 'causam', 'irritação' e 'intestino' numa relação de colocação conduzem o leitor à identificação de que algum problema está acontecendo no aparelho digestivo, alertando-o para a procura de informações relacionadas a esse problema. As palavras 'infecções', 'gastroenterites' e 'medicamentos' auxiliam o leitor nessa procura e preenchem informações relevantes, ou seja, preenchem a lacuna de causadores do problema.

Além do verbo 'causar', outros verbos causativos, tais como 'perfurar', também poderiam indicar o mesmo tipo de relação causa e efeito, como na ocorrência (33):

33) [...] o primeiro sintoma de uma *úlcer*a péptica pode ser a presença de *sangue nas fezes*. Quando uma *úlcer*a perfura o estômago ou o **intestino** delgado, a criança pode demonstrar *sinais de dor* [...]

A relação de causa e efeito em (33) se evidencia a partir das combinações das palavras 'úlcer'a', 'perfurar', 'intestino' e 'sinais de dor'. A ativação de traços semânticos

negativos presentes nas expressões ‘sangue nas fezes’ e ‘sinais de dor’, permite ao leitor concluir que algum problema está acontecendo no ‘intestino’. O que inicialmente poderia ser somente uma suposição para o leitor, passa a ser uma confirmação quando a palavra ‘úlcera’ se apresenta.

Importa mencionar que a interpretação de causa e efeito fica evidenciada nas ocorrências registradas acima, pela co-ocorrência com verbos com um traço semântico ativo e causativo (‘causar’, ‘perfurar’) que fornecem à palavra ‘intestino’ um caráter de passividade, configurando-se como receptor de um processo negativo, seja ele um sintoma (irritação em 64) ou uma patologia (úlcera em 33). Em outro exemplo,

82) Quando a doença de Crohn ou colite ulcerativa *afeta* o **intestino** grosso, esse processo é alterado, *resultando em sintomas* de diarreia, urgência na evacuação, sangramento retal e dor abdominal.

a coligação com o verbo causativo ‘afetar’ e a colocação com palavras negativas como ‘doença’ e ‘sintomas’ também confirmam a indicação de uma relação de causa e efeito, e ‘intestino’ aparece como o receptor dessa relação. Nessa oração, há ainda a locução verbal ‘resultar em’, que alerta o leitor à direta extração de informações relativas aos sintomas da patologia no organismo.

Segundo Kleiman (2004), na maioria das vezes, as relações causativas são consideradas temáticas, refletindo a estrutura referencial do texto expositivo. Nos exemplos acima expostos, percebe-se claramente que a ocorrência da relação de causa e efeito na ambiência de ‘intestino’ propicia uma interpretação relativa a um possível problema (patologia, sintoma) e seus efeitos no organismo, constituindo dois tipos de informações essenciais para interpretação dos textos de estudo.

4.3.5 'Intestino' adotando um papel ativo no sistema digestivo.

Nas categorias analisadas até o momento, a palavra 'intestino', na maioria das vezes, apareceu com um traço semântico passivo. No entanto, nem sempre essa palavra assume essa característica: ela também pode assumir um sentido ativo, desde que entendida como a denominação de um órgão com funções específicas no sistema digestivo. Tome-se por base o exemplo (55):

55) [...] 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' estabelece na leitura uma expectativa para a aparição das atribuições do órgão em questão, lacuna essa preenchida pelos verbos 'reabsorver' e 'eliminar', que indicam atividade para 'intestino'. Note-se também o exemplo (61):

61) 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 [...]

Como no caso anterior, o verbo 'desagregar' também atribui à palavra 'intestino' uma função ativa. Por meio da percepção da palavra-chave 'intestino' em co-ocorrência com palavras como os verbos ativos 'desagregar', 'reabsorver' e 'eliminar', é possível observar que a interpretação dessas porções refere-se às funções atributivas do

‘intestino’, e não a patologias, como pudemos constatar em algumas das categorias anteriores.

4.3.6 Extração de informação interpretativa a partir dos padrões lexicais da palavra ‘intestino’.

Embora provavelmente haja outras manifestações relevantes a partir da palavra ‘intestino’ e seus colocados e coligados, os exemplos apresentados anteriormente já seriam suficientes para demonstrar que esta palavra-chave (assim como as outras palavras-chaves que serão comentadas na seqüência) revela em sua ambiência situações semânticas que conduzem o leitor a estabelecer previsões e levantar hipóteses muito sugestivas para o processo de interpretação. Por exemplo, revisemos o caso em que ‘intestino’ aparece como locativo-receptor ou como órgão de passagem. Das análises feitas, foi possível distinguir ao redor de ‘intestino’ palavras que indicam um sentido de movimento e direção. Nessa relação, informações interpretativas sobre o direcionamento dos alimentos no processo digestivo ou sobre o funcionamento do mesmo podem ser extraídas das ocorrências. O trecho a seguir servirá para nos dar uma melhor idéia sobre isso:

(51) a (55) 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. 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.

Nesta porção do texto, a interpretação é conduzida para algo que enfoque o funcionamento do processo digestivo. Tal interpretação é possibilitada pela relação léxico-semântica mantida pela coligação entre a palavra-chave ‘intestino’ e locuções verbais que indicam passagem e localização, como ‘passar por’ e ‘é completada em’, as quais estabelecem para a sentença a idéia de processo. A palavra associada ‘digestão’ também contribui para a ativação dessa interpretação. Neste caso, a partir dos traços semânticos de movimento, processo e passagem, pode-se retirar uma informação do tipo “o alimento passa por várias etapas durante a digestão”.

Ainda conforme vimos na análise, descreveu-se para a palavra ‘intestino’ o papel semântico (4.3.5) em que há uma interpretação das funções deste órgão, a qual é estabelecida a partir do papel ativo assumido pela palavra-chave. Veja-se, no exemplo (55) acima, uma coligação com verbos com traço semântico ativo (‘reabsorver’, ‘eliminar’), estabelecendo para o parágrafo uma possibilidade interpretativa conectada às atribuições específicas deste órgão no sistema digestivo. A informação relativa a “o que o intestino faz no organismo?” está claramente estabelecida aqui pela coligação “intestino-verbo ativo”.

Outro caso ainda foi observado no papel semântico indicativo de localização dos órgãos do sistema digestivo, como no exemplo (38) em 4.1.3. Colocações importantes com substantivos tais como ‘duodeno’, ‘estômago’, ‘esôfago’ e, principalmente, com ‘exames’, fornecem à ocorrência em contexto os subsídios interpretativos necessários para a identificação de que há um exame diagnóstico a ser realizado:

(38) A *endoscopia* é um procedimento que permite o *exame* do revestimento interno do *esôfago*, *estômago* e *duodeno* (1ª porção do **intestino** fino). O endoscópio é um aparelho flexível, de diâmetro menor que 1 cm, colocado através da boca e dentro do esôfago, estômago e duodeno.

As relações diretas estabelecidas entre a palavra-chave e seus colocados (duodeno, estômago, esôfago), e as palavras ‘exame’ e ‘endoscopia’, levam o leitor a ativar esquemas semânticos relacionados a exames, o que possibilita extrair desta porção a informação: “O exame da endoscopia é realizado para diagnosticar patologias gástricas.”

Também quando ‘intestino’ assume o papel de receptor na relação de causa e efeito, informações importantes podem ser extraídas para a interpretação dos textos a partir dos padrões lexicais da palavra-chave. Seja o exemplo:

(33) 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.

As colocações com palavras com carga semântica altamente negativa (por exemplo, ‘úlcera’) e coligações com verbos causativos (por exemplo, ‘perfurar’) estabelecem nesta ocorrência a expectativa de uma interpretação voltada aos efeitos da patologia em questão. Neste extrato ainda, a aparição dos sintomas ‘sinais de dor’ e ‘sangue nas fezes’ fornece ao leitor as informações necessárias para o preenchimento dessa expectativa.

As colocações e coligações de ‘intestino’, associadas à prosódia semântica negativa (ou em alguns casos neutra) estabelecem para as ocorrências de ‘intestino’ a possibilidade de se extrair informações essenciais para a interpretação do conteúdo dos textos, tais como “De quais sintomas fala ao texto?”; “Qual patologia é descrita?”;

“O que acontece no processo digestivo?”. Esta indicação do conteúdo proposicional também se mostrou na análise das demais palavras-chaves.

4.4 As relações contextuais da palavra ‘causa’

A exemplo da palavra ‘intestino’, as outras palavras-chaves rastreadas para esse estudo também foram analisadas a partir dos padrões lexicais, a fim de se retirar delas e da sua ambiência subsídios para a interpretação dos textos. Com relação à palavra ‘causa’, 108 ocorrências foram analisadas, cuja distribuição nos textos é resumida na tabela abaixo:

Tabela III: 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 Adjetivos (formando expressões: causa funcional, causa psicológica) 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 interpretação sobre 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. Alguns colocados são mais freqüentes neste tipo de ocorrência, tais como ‘úlceras’, ‘gastrite’ e ‘doença’, indicando a possibilidade dos textos tratarem de patologias 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 da coligação com verbo de ligação.

Na manifestação de ‘causa’, há ainda uma coligação com adjetivos formando uma expressão (‘causa funcional’, ‘causa psicológica’) de modo a restringir o significado da palavra-chave. Tome-se como amostra o exemplo (31):

(31) [...] a *dor* abdominal recorrente tem uma **causa psicológica** e *não* uma **causa funcional**. A dor de **causa psicológica** aparentemente é desencadeada ou piorada pela tensão, ansiedade ou depressão.

Ao deparar-se com a palavra ‘dor’, o leitor automaticamente gera uma expectativa de encontrar ali informações relacionadas aos sintomas de uma determinada doença. O adjetivo ‘psicológica’ e a negação anterior à palavra ‘funcional’ quebram essa expectativa, fazendo com que o leitor reconstrua suas informações de tal

maneira que a interpretação dada a essa porção textual seja, de fato, sobre um sintoma, porém sem ligação à patologia aparente.

Há um grande número de ocorrências da palavra ‘causa’ com colocados que indicam sintomas (‘dor’) e patologias (‘dispepsia’, ‘gastrite’, ‘úlcera’, ‘doenças’), estabelecendo para o ambiente lingüístico a expectativa da aparição da descrição de tais assuntos. Na seqüência, serão mostradas algumas possíveis distribuições que puderam ser identificadas, geradas pela palavra ‘causa’ em diferentes contextos.

4.4.1 ‘Causa’ como indicadora do causador de um problema

Com base em alguns colocados da palavra ‘causa’ com a função de verbo, é possível perceber a presença de um causador (agente) desencadeador de um acontecimento. O exemplo abaixo nos fornece subsídios para essa afirmação:

11) [...] 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 (11), as expressões ‘rompimento’ e ‘hemorragia digestiva’ estabelecem uma conotação semântica negativa que leva o leitor à identificação do 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 = hemorragia digestiva. Veja-se em mais um exemplo:

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

Similarmente ao exemplo (11), em (101) é possível evidenciar as informações relacionadas ao causador e ao acontecimento. A palavra ‘solução’ identifica imediatamente o causador de algum tipo de irregularidade no sistema digestivo, e a palavra ‘desconforto’ indica o sintoma por ele desencadeado.

Nessa sua realização verbal, o que se observa é uma relação direta, na qual as informações, como patologias e sintomas, ficam mais evidentes através dos próprios argumentos do verbo ‘causar’:

30) A maioria dos *distúrbios gastrointestinais* **causa** dor. Muitos deles (p.ex., doença celíaca e a intolerância à lactose) também causam desnutrição e diarreia.

Agente / causador (argumento 1): distúrbios gastrointestinais

Acontecimento / efeito (argumento 2): dor (sintoma)

Neste exemplo, identificam-se os argumentos 1 e 2, e a ativação dos traços semânticos dos mesmos propicia a identificação do assunto dos textos: problemas do sistema digestivo e seus sintomas. No exemplo (30) acima, outras patologias e seus sintomas, tais como “doença celíaca = desnutrição”, podem ainda ser percebidas. Essa relação causal também é notada no exemplo (81):

81) [...] 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 caso, a palavra 'doença de Chron' leva o leitor a identificar a patologia descrita, e, na seqüência da palavra 'causa', o leitor identifica os sintomas a ela relacionados.

Destaca-se que a palavra 'causa' mantém em torno de si um ambiente que revela o assunto tratado nos textos. Com base na inter-relação dos itens lexicais e das ativações de seus significados, o leitor, aos poucos, estabelece o assunto. 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 (CAVALCANTI, 1989). Esta 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 nos fornecer subsídios importantes relacionados ao conteúdo principal do texto, como, por exemplo, 'doença de Chron causa inflamação, ulceração e estenoses', em (81).

Os exemplos abaixo nos fornecem uma idéia mais precisa:

19) Nestes pacientes só há indicação de cirurgia quando a *hérnia* é muito *grande* e **causa** compressão de órgãos próximos a ela como pulmão e coração.

Agente / causador (argumento 1): hérnia grande

Acontecimento / efeito (argumento 2): compressão de órgãos

Reiteramos aqui que os argumentos do verbo, colocados naturais da palavra-chave 'causa', fornecem dados essenciais para a interpretação dos textos. No exemplo (19) acima, o argumento 'hérnia grande' revela a patologia, e a expressão 'compressão dos órgãos' denota o efeito no organismo. No exemplo (27),

27) A *úlcera* geralmente **causa** dor e queimação na parte superior do abdômen.

Agente / causador (argumento 1): *úlcera*

Acontecimento / efeito (argumento 2): dor e queimação (sintomas)

a extração dos argumentos 'úlcera' e 'dor e queimação' estabelece, respectivamente, o sentido sobre causador (patologia) e efeito (sintomas). Esses dados são altamente relevantes para a extração de informações relacionadas à doença e ao assunto geral do texto.

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:

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

18) [...] o que **causa**, então, o *refluxo*?

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, criando algumas lacunas interpretativas que deverão ser mais facilmente preenchidas no decorrer da leitura do texto. De acordo com Kleiman (2004: 111),

uma seqüência interrogativa é uma pergunta-retórica porque o contexto total indica que não haverá oportunidade para o interlocutor responder; daí que sabendo que a informação se constrói no par pergunta-resposta, esperamos uma resposta fornecida pelo próprio autor, e a procuramos durante a leitura.

Verificou-se, também, que ocorrências tais como (21) e (18) 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 confirma a importância da palavra-chave 'causa' como indício interpretativo para os textos.

O verbo 'causar' parece apoiar-se na semântica de seus argumentos para estabelecer sentido nas sentenças, quase como numa relação de anáfora ou catáfora. Porém, essa relação fica mais nítida na ambiência do substantivo 'causa', como veremos na seqüência.

4.4.2 A palavra 'causa' como indicadora da conseqüência de um problema.

Uma das ocorrências de 'causa' como substantivo freqüentemente evidencia uma relação de conseqüência. Semelhantemente à sua realização verbal, evidencia-se um problema e imediatamente a conseqüência disso no organismo. Veja-se o exemplo a seguir:

29) [...] 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: 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 ‘úlceras’ revela a patologia descrita no texto. Para o leitor que procura no texto informações relativas a uma patologia, a identificação dessas combinações não pode passar despercebida no processo interpretativo.

Mais um indicativo da relação de conseqüência é a coligação (vide tabela 2) em que aparece um verbo de ligação (verbo de ligação+causa+de), como mostra o exemplo (17) abaixo:

17) [...] Esse tipo de medicação, aparentemente simples, amplamente usada e vendida sob diversos nomes comerciais é uma **causa** importante e grave de *hemorragia digestiva* que pode ser evitada [...]

Problema: uso de medicamentos

Conseqüência: hemorragia digestiva.

A coligação mencionada funciona como um marcador do tema principal abordado no texto. Em (17), por exemplo, “a medicação é causa de hemorragia digestiva” fornece ao leitor uma previsão de que o assunto deste extrato é sobre o sintoma ‘hemorragia’ e o leva a procurar maiores informações sobre o assunto, facilitando o seu trabalho de interpretação e extração de informação. Na ocorrência abaixo,

74) [...] 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: câncer colo-retal

Consequência: morte

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 fazem 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 proficientes reagem à organização hierárquica das informações, mostrando grande convergência naquilo que consideram mais importante (...) sendo que o que é percebido como mais importante é a informação que depende diretamente do tópico.

Em outros termos, 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 na primeira dessas tarefas.

Outro fato que chama a atenção é que, da mesma forma que acontecia com o verbo 'causar', percebe-se a aparição desta palavra-chave no título ou em resposta a

perguntas, porém na forma de substantivo. O exemplo abaixo nos auxilia nesta percepção:

75) 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.

Reforçamos 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 como a anáfora ou a catáfora fornece à 'causa' o sentido necessário para sua interpretação, conforme se pode verificar nos exemplos:

12) [...] podem ser surpreendidas por eventual *vômito com sangue vivo*. A **causa** disto são as chamadas *fissuras de Mallory Weiss*, que são pequenas [...]

No exemplo (12), o pronome 'disto' logo na seqüência da palavra 'causa' mantém uma relação anafórica com "vômito com sangue vivo", estabelecendo para a palavra-chave em análise a interpretação negativa. Tal expressão, de certa forma, insere-se na palavra 'causa' como um meio de integrar informações e sobrepor participantes.

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. Por exemplo, no caso da sentença (12), ele localiza o pronome 'disto' em relação anafórica com 'vômito com sangue vivo' e integra isso à informação nova sobre a patologia: fissuras de Mallory Weiss.

Analisemos mais uma ocorrência:

24) *Helicobacter pylori* é o nome que identifica a **causa** da maioria das *úlceras*, excluindo aquelas causadas por aspirina ou antiinflamatórios [...]

Em (24), um sentido altamente negativo relacionado à patologia é percebido por meio de uma colocação com '*Helicobacter Pylori*' ao redor de 'causa', a qual mantém uma relação catafórica com 'úlceras'. Neste exemplo, o leitor procura na catáfora o sentido a que se relaciona a palavra 'causa'. A palavra 'úlceras' completa a informação exigida pelo substantivo: causa de quê? Assim, nas relações catafóricas e anafóricas, o leitor estabelece as conexões necessárias para preencher o sentido da palavra 'causa' nos textos, reconhecendo algumas das interpretações necessárias para cada sentença.

4.4.3 Extração de informação interpretativa a partir dos padrões lexicais da colocação, coligação e prosódia semântica na palavra 'causa'.

Semelhanamente à primeira palavra-chave analisada, é possível observar, por meio do ambiente semântico formado pela palavra ‘causa’, algumas possibilidades interpretativas. Veja-se, por exemplo, a relação causal discutida em 4.2.1, que determina um acontecimento relacionado a uma patologia:

(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’.

Seria muito comum o surgimento de um questionamento: “o que significa a palavra ‘causa’?”. Conforme discutido anteriormente, sozinha esta palavra não assume sentido algum. Ela necessita se apropriar da carga semântica dos elementos que com ela se colocam ou coligam nos textos. As informações semânticas desses elementos “enxertadas” no conteúdo lexical da palavra ‘causa’ é que instituem indícios importantes para a interpretação dos textos.

Para Kato (1999), o item não aparece isolado no léxico mental, mas no âmbito de uma rede de relações paradigmáticas e sintagmáticas. Assim, o aparecimento de um determinado item no texto deveria ativar outros itens da mesma área, levando a um

desencadeamento temático do texto. No caso dos textos de estudo, palavras tais como ‘úlceras’, ‘problema’, ‘doença’, etc., conduzem o leitor a fazer previsões acerca de um possível assunto. Sem essa previsão, a interpretação se tornaria, no mínimo, difícil. A palavra ‘causa’, nestes casos, aparece como vinculadora do tema e como indicadora importante das patologias, sintomas e suas descrições.

Veja-se o exemplo (18), em que a palavra ‘causa’ aparece no subtítulo do texto. Conforme destacamos em 4.1.2, a pergunta em que a palavra-chave é expressa faz com que o leitor naturalmente tenha de construir lacunas interpretativas, as quais devem ser preenchidas quando ele localiza a informação necessária durante o processo de leitura. Note-se o contexto da sentença:

(18) O que **causa**, então, o *refluxo*? Resumidamente, ele é resultado do *mau funcionamento da válvula* que fica na parte superior do estômago. É como se esse músculo estivesse com problemas de regulação, permitindo o retorno do alimento.

Uma lacuna interpretativa do tipo: “_____ causa refluxo” é imediatamente formada a partir da aparição do item ‘causa’ na pergunta. Na leitura da seqüência do texto, o leitor pode facilmente destacar “mau funcionamento da válvula” para preencher tal lacuna e extrair a informação que procura.

Outra possível interpretação ligada à palavra ‘causa’ é revelada pela relação de conseqüência de um determinado problema no organismo. Por exemplo, na ocorrência (29), a informação “*helicobacter pylori* é causa de úlceras”, percebida pelas colocações que indicam o problema e a conseqüência, institui claramente a interpretação relacionada à patologia e a seus sintomas, como mostra o contexto:

(29) Descobriu-se 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 antiinflamatórios não hormonais contribuem para o aumento excessivo do ácido clorídrico no estômago, o que também pode causar a doença.

A descrição dos sintomas e patologias estabelece a prosódia negativa, a qual suscita um estado de alerta no leitor. Suponhamos que o leitor procure informações úteis para sua própria saúde. Ele interpreta as informações relacionadas ao assunto 'úlcera', destacando informações do tipo: "a úlcera pode ser causada pela bactéria", ou "fumo, estresse e medicamentos podem causar a úlcera". Destaca-se que o perfil dos textos de estudo está relacionado à informação de pessoas "leigas", que procuram nos textos maiores informações sobre os problemas, sintomas, causas, etc. de patologias digestivas. O ambiente lingüístico formado pelas palavras-chaves revela essas informações, assim como o assunto da maioria dos textos, como é o caso do exemplo (12):

12) Certas pessoas que vomitam muito, como por exemplo, aquelas que bebem demais, podem ser surpreendidas por eventual *vômito com sangue vivo*. A **causa disto** são as chamadas *fissuras de Mallory-Weiss*, que são pequenas lacerações (rachaduras) longitudinais da mucosa da junção do esôfago com o estômago, causado pela excessiva pressão nesses órgãos, durante o vômito.

Na análise, destacou-se a relação direta da palavra-chave 'causa' com os colocados 'vômito com sangue vivo' e 'fissuras de Mallory-Weiss'. Esta patologia não constitui algo comum no vocabulário diário de pessoas que não trabalham na área gástrica. Um leitor que procura no texto informações relativas a essa patologia provavelmente destacaria informações relacionadas a: "o que é?", "quais os sintomas?". A primeira dessas questões é facilmente respondida ao redor de 'causa'

por meio de colocados na oração subordinada que preenche os traços semânticos da doença: ‘pequenas lacerações (rachaduras)’ e ‘da mucosa’. E o colocado ‘vômito com sangue vivo’ preenche nitidamente a segunda das questões.

As relações da palavra ‘causa’ revelam claramente possibilidades interpretativas altamente relevantes para os textos de estudo. Por meio dos padrões lexicais, o leitor é levado a estabelecer previsões importantíssimas para a interpretação dos textos do tipo: “este texto descreve os efeitos da úlcera”, ou ainda, “este texto trata de estabelecer quais são as causas de determinada patologia”.

4.5 As relações contextuais da palavra ‘*Helicobacter Pylori*’

Outra palavra superchave selecionada para esse estudo foi ‘*Helicobacter pylori*’. A exemplo das primeiras palavras analisadas, tomou-se por base um total de 89 ocorrências observadas em seu contexto. As ocorrências dessa palavra podem ser resumidamente examinadas na tabela a seguir:

Tabela IV: Lista das colocações, coligações e prosódia semântica mantidas pelo item ‘*helicobacter pylori*’

OCORRÊNCIAS DA PALAVRA ‘HELICOBACTER PYLORI (HP)’		
COLOCAÇÕES	COLIGAÇÕES	PROSÓDIA SEMÂNTICA PREDOMINANTE

Bactéria (22) Gastrite (21) Infecção (17) Causada (12) Crônica (9) Erradicação (7) causa (5)	Substantivos HP+predicado nominal (Clusters: a infecção causada pelo, gastrite causada, uma bactéria)	Negativa
--	---	----------

Alguns colocados de ‘*Helicobacter pylori*’ (HP), já fornecem algumas pistas sobre quais papéis essa palavra pode assumir nos textos de estudo. Veja-se na tabela, por exemplo, um dos colocados mais freqüentes: ‘bactéria’. Essa palavra é uma associada, ou seja, também é observada como palavra-chave nos textos e tem íntima relação com HP.¹⁵ A palavra associada, neste caso, aparece claramente para auxiliar o leitor no processo de conceituação de ‘HP’, ou seja, ela aparece para preencher traços semânticos da palavra-chave.

A exemplo da palavra ‘causa’, colocados tais como ‘gastrite’ e ‘infecção’ já levam à ativação de uma previsão interpretativa referente a patologias. O exemplo abaixo confirma isso:

61) A inflamação da mucosa do duodeno pode aparecer associada à *gastrite causada* pelo **H. pylori**. Essa mucosa inflamada do bulbo duodenal pode ser substituída por mucosa do estômago [...]

O colocado ‘gastrite’ indica, na sentença, a referência a uma patologia. O verbo ‘causada’, na sentença em voz passiva, fornece à palavra-chave a responsabilidade

¹⁵ Na seção 4.1 deste trabalho, discutimos o motivo pelo qual estamos analisando a palavra ‘bactéria’ em ligação com ‘*helicobacter pylori*’.

pela manifestação da doença. Essas observações já nos levam a prever uma possibilidade de interpretação voltada à descrição da doença.

Em (34) abaixo, também se observa a relação de colocações que fornecem a 'HP' papéis semânticos claros:

34) A grande *vilã* da história é *uma bactéria*, a ***Helicobacter pylori***, presente em todos os casos de *gastrite*.

Neste caso, há uma relação nítida entre as palavras 'bactéria' e 'gastrite'. A carga negativa assumida por esses colocados é reforçada pela aparição da palavra 'vilã', o que faz com que o leitor automaticamente atribua uma carga igualmente negativa ao conteúdo semântico de 'HP'.

A relação forte entre palavras tais como 'HP', 'gastrite' e 'infecção' confirma-se pela observação de alguns *clusters*¹⁶ como “a infecção causada pelo”, “gastrite causada”, “uma bactéria”.

Há também o colocado 'erradicação', que, assim como as outras relações de colocação, conduz a uma alta negatividade do teor semântico da palavra-chave. Deve-se lembrar que, diferentemente das outras palavras-chaves analisadas, 'causa' e 'intestino', a palavra 'HP' já possui em suas características semânticas uma conotação negativa. Porém, essa carga é reforçada nos textos por seus co-ocorrentes também negativos, como 'bactéria', 'gastrite', 'infecção' e 'erradicação'.

Embora a constituição semântica desta palavra-chave seja, de certa forma, diferente das palavras analisadas anteriormente, também para ela foi possível destacar

¹⁶ *Clusters* são palavras encontradas repetidamente em companhia uma da outra e na sequência. Elas representam uma relação mais forte que colocados, como unidades de várias palavras ou grupos de palavras. (SCOTT, 2004) (Do original: “*Clusters are words which are found repeatedly together in each others' company, in sequence. They represent a tighter relationship than collocates, more like multi-word units or groups*”).

algumas possibilidades interpretativas que foram levantadas a partir dos padrões lexicais de 'HP'.

4.5.1 O ambiente lingüístico de 'HP' estabelecendo conceito.

Os colocados que aparecem com mais freqüência com a palavra-chave 'HP', tais como a palavra associada 'bactéria', contribuem para a sua conceituação. Uma vez que 'HP' não é uma palavra comum a muitas pessoas, e, algumas vezes, seu conceito precisa ser construído no léxico mental durante a leitura, tais informações nos fornecem os traços semânticos necessários para a palavra, como no exemplo:

64) [...] sabemos desde 1983 que a causa mais freqüente de *gastrite crônica* é uma *bactéria*: o [Helicobacter pylori](#). A *gastrite crônica* causada pelo **Helicobacter pylori** (também designada por gastrite B) [...]

Em (64), há uma evidente relação lexical mantida entre a palavra-chave 'HP' e os seus colocados 'bactéria' e 'gastrite'. A primeira dessas palavras preenche o significado de 'HP', e a segunda revela uma relação em que 'HP' aparece como causador da patologia (gastrite). O verbo 'causada', na voz passiva, ajuda a estabelecer essa relação. Esta observação fornece informações que auxiliam a interpretação do tipo: HP+conceito+patologia. Tomemos como exemplo as ocorrências (1) e (10) abaixo:

1) **Helicobacter Pylori**. Esta é *uma bactéria* que está presente em aproximadamente 70 % na população brasileira. É atribuída a esta *bactéria* a *formação da gastrite* [...]

Neste caso, a relação entre a palavra ‘bactéria’ e a expressão ‘formação da gastrite’ indica ‘HP’ como responsável pela manifestação da patologia. O verbo ‘atribuir a’ na voz passiva intensifica essa relação.

10) [...] hoje sabe-se que o *Helicobacter Pylori* é o *causador de perturbações* na mucosa gástrica. [...]

No exemplo (10), um traço semântico de ‘HP’ é dado pela expressão ‘causador de perturbações’, que fornece à palavra-chave em análise um papel semântico ativo e maléfico ao organismo. Nos dois exemplos acima, a coligação dada entre HP+predicado nominal indica um sintagma que fornece um conceito a ‘HP’, como é o caso de ‘uma bactéria’ em (1), e ‘causador de perturbações’ em (10).

Esta relação de ‘HP’ com seu conceito não é a única possibilidade para a interpretação das porções textuais em que esta palavra-chave aparece. Constatou-se, ainda, em algumas ocorrências, a relação direta de ‘HP’ com a manifestação não só da gastrite, mas de diferentes patologias gástricas.

4.5.2 A indicação lingüística da palavra ‘HP’ como causador de patologias.

Informações relacionadas ao que a bactéria ‘HP’ causa no organismo são dadas por palavras que sugerem que ‘*Helicobacter Pylori*’ é um agente causador de patologias do sistema digestivo, como é o caso da gastrite em (11), e da úlcera em (26):

11) [...] A gastrite pode ser causada por ingerir muita bebida alcoólica, uso prolongado de remédios anti-inflamatórios não-esteróides, ou *infecção por bactéria* como a

Helicobacter pylori. Algumas vezes a *gastrite* desenvolve-se depois de uma grande cirurgia [...]

Neste exemplo, percebe-se que, além da ocorrência direta com a palavra ‘bactéria’, que contribui para o seu conceito, ‘HP’ também mantém uma relação clara com problemas gástricos como ‘infecção’ e ‘gastrite’.

Veja-se, ainda, o exemplo (26) a seguir:

26) [...] Descobriu-se que uma *bactéria* chamada **Helicobacter Pylori** é a grande responsável pela causa das *úlceras* [...]

Similarmente ao exemplo (11), em (26) ‘HP’ mantém a relação conceitual com a palavra ‘bactéria’. No entanto, outra patologia gástrica, a ‘úlcera’, é ativada no ambiente lingüístico da palavra. Destaca-se que em ambas as sentenças, (11) e (26), é possível estabelecer a interpretação “conceito+patologia” com base no item lexical nuclear ‘HP’ e seus colocados: ‘bactéria’, ‘gastrite’, ‘úlcera’. Em outros termos, a partir das colocações em (11) – ‘infecções’, ‘bactérias’, ‘HP’, ‘gastrite’ – é possível extrair a informação: infecção por bactéria HP = gastrite. O mesmo caso aplicado à sentença (26) também fornece uma informação importante: bactéria HP = úlcera.

Dado que uma das características principais dos textos de pesquisa é proporcionar dados sobre a causa de determinados problemas gástricos, o ambiente lingüístico formado pela palavra-chave ‘HP’ parece fornecer indícios essenciais para a extração dessa informação. ‘HP’ aparece como responsável por várias das patologias descritas no *corpus* de pesquisa.

4.5.3 A relação lingüística entre ‘HP’ e uma necessidade de eliminação.

Conforme vimos até o presente momento da análise, a palavra ‘HP’ associa-se com colocados com carga semântica altamente negativa. A interpretação de ‘HP’ como causador de patologias como a ‘úlcera’ e a ‘gastrite’ levam o leitor a inferir que a bactéria traz grandes malefícios ao organismo. A prosódia semântica negativa de ‘HP’ é ainda reforçada por palavras que instituem a necessidade de exclusão da bactéria do organismo, como indicam os exemplos abaixo:

15) Sabe-se que a *eliminação* do ***Helicobacter Pylori*** com o *uso de antibióticos* associados a inibidores da secreção ácida resulta, na maioria dos ulcerosos, na *cura* da doença [...]

75) [...] o objetivo do *tratamento da gastrite* causada pelo ***Helicobacter Pylori*** é a *erradicação* do microorganismo que é definida como a negatificação dos exames quatro semanas após o fim do uso de antimicrobianos [...]

Em (15) e (75), as palavras ‘erradicação’ e ‘eliminação’, respectivamente, reforçam em ‘HP’ uma carga negativa. Se algo precisa ser erradicado ou eliminado, semanticamente constitui-se como algo maléfico. Essa relação de ‘HP’ com sua eliminação constitui a informação relativa ao tratamento da doença, o que pode ser observado pela aparição da expressão ‘uso de antibióticos’ e da palavra ‘cura’ em (15), e das expressões ‘tratamento da gastrite’ e ‘negatificação dos exames’ em (75).

A utilização da forma verbal ‘erradicar’ também institui a interpretação de eliminação:

76) [...] Há *consenso internacional* quanto à necessidade de *erradicar* o ***Helicobacter Pylori*** quando ele coexiste com *doença ulcerosa*, mas em relação a outras situações clínicas como a gastrite, existem ainda controvérsias [...]

O verbo ‘erradicar’ e a expressão ‘doença ulcerosa’ levam o leitor a estabelecer para ‘HP’ um traço semântico negativo e não pertencente ao organismo humano. Expressões como ‘consenso internacional’ e ‘controvérsias’, em (76), nos remetem à noção de que ‘HP’ é um foco de discussões na área gástrica. Esse sentido é confirmado também por palavras indicadoras de pesquisas, como veremos na seqüência.

4.5.4 A relação entre ‘HP’ e palavras indicadoras de pesquisas.

A palavra ‘bactéria’, conforme discutimos anteriormente, aparece como uma âncora para o entendimento de ‘HP’ nos textos. Esse processo de conceituação parece normal, pois mesmo para especialistas da área gástrica, ‘HP’ é um microorganismo ainda em estudos. As colocações de ‘HP’ com palavras ligadas ao sentido de pesquisas demonstram isso. Elas evidenciam a descoberta recente e os estudos relacionados à bactéria. O exemplo abaixo fornece uma idéia mais clara:

36) Em 1987, a *equipe* desenvolveu um novo meio de cultura da ***Helicobacter Pylori***, que é utilizado por *pesquisadores* de todo o mundo.

A colocação de ‘HP’ com ‘pesquisadores’ e ‘equipe’ institui para a sentença uma possível interpretação voltada aos estudos realizados sobre a bactéria. Ainda, o ano

‘1987’ fornece dados relativos à recente descoberta do ‘HP’. Esse sentido também pode ser observado em:

89) *A recente descoberta da estrutura genética (genoma) do **Helicobacter Pylori** permitirá o desenvolvimento de estratégias para sua erradicação [...]*

Em (89), a expressão ‘descoberta do genoma’ fornece ao núcleo ‘HP’ traços semânticos relacionados a pesquisas. A própria palavra ‘genoma’, por ser uma palavra de formação recente na área científica, institui na sentença o sentido de novidade, o que é ainda reforçado pela palavra ‘recente’, que contém traços semânticos similares.

A partir dessas ocorrências, é possível sugerir que ‘HP’ é um assunto gerador de polêmica na área de gastroenterologia. A percepção de que, de fato, a bactéria afigura-se como agente causador não de uma, mas sim de várias patologias em ligação, correlacionada ao fato de haver uma necessidade de pesquisas, pode levar o leitor a estabelecer previsões importantes e a procurar informações relativas às possíveis conseqüências da bactéria no organismo. Há a possibilidade, ainda, de o leitor identificar várias das patologias descritas nos textos.

4.5.5 Extração de informação interpretativa a partir dos padrões lexicais da palavra ‘*Helicobacter Pylori*’.

A exemplo das palavras-chaves analisadas anteriormente, a palavra ‘HP’ também revela possibilidades interpretativas interessantes com base em seus padrões e na carga semântica altamente negativa. Esta palavra apresenta uma relação como

agente causador de diferentes patologias gástricas, conforme demonstrado no exemplo em contexto abaixo:

(26) Descobriu-se 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 antiinflamatórios não hormonais contribuem para o aumento excessivo do ácido clorídrico no estômago, o que também pode causar a doença.

Neste trecho, percebe-se a clara interpretação de que ‘HP’ é um dos principais causadores da úlcera, o que se configura pela relação direta assumida entre as palavras ‘bactéria’, que estabelece o conceito, ‘úlcera’ que indica a patologia sendo descrita, e ‘responsável’, que fornece à ‘HP’ a carga negativa de agente causador. Na seqüência, a palavra ‘fatores’, o verbo ‘contribuir’ e a expressão ‘causar a doença’ fazem com que o leitor volte sua atenção para a retirada da informação: “quais são algumas das possíveis causas da úlcera?”.

O papel semântico assumido por ‘HP’ como causador de patologias e a carga conceitual estabelecida por colocações como ‘bactéria’ e ‘causador’, associados à necessidade de eliminação evidenciada por palavras tais como ‘erradicação’, fornecem ao leitor alguns subsídios interpretativos de que a cura da doença implica necessariamente a expulsão da bactéria. Veja-se o contexto do exemplo (15):

(15) Sabe-se que a *eliminação* do **Helicobacter pylori** com o *uso de antibióticos* associados a *inibidores* da secreção ácida resulta, na maioria dos ulcerosos, na *cura* da doença e que, nas pessoas nas quais a *bactéria é eliminada*, a taxa de recidiva, isto é, o retorno da úlcera é insignificante.

Nesta porção de texto, a palavra ‘eliminação’ em associação com ‘uso de antibióticos’ e ‘inibidores’ fornecem ao leitor a informação relativa a tratamento. O aparecimento da palavra ‘cura’ reforça a idéia de que a solução para a patologia causada por ‘HP’ é, de fato, eliminá-la do organismo. Essas interpretações, relacionadas à expulsão da bactéria do organismo, aparecem associadas a trechos que indicam para ‘HP’ a relação com pesquisas. Veja-se o extrato abaixo:

(89) A recente descoberta da *estrutura genética (genoma)* do ***Helicobacter Pylori*** permitirá o desenvolvimento de estratégias para sua *erradicação* seja através de vacinas ou da melhor ação dos antibióticos contra esse microorganismo.

A necessidade de ‘erradicação’ associada a um sentido relacionado a estudos, dado por palavras tais como ‘descoberta’ e ‘genoma’, institui para as ocorrências analisadas a interpretação voltada à exposição de um problema ainda em estudo na área médica. Destacamos, também, no exemplo (26) apresentado acima, a aparição da palavra ‘descobriu-se’, que, mesmo em um contexto ligeiramente diferente, confirma a interpretação de pesquisas.

Em suma, pode-se extrair destas relações de ‘HP’ a informação de que a bactéria mantém uma relação estreita como causadora de diversas patologias gástricas descritas no decorrer de todos os textos do *corpus* de estudo, entre elas, infecções, gastrite e úlcera. A exemplo das primeiras palavras analisadas, tal observação evidencia a importância desta palavra-chave na identificação do assunto dos textos e no reconhecimento de informações importantes, tais como os sintomas e as causas das patologias.

5 CONSIDERAÇÕES FINAIS

Como vimos no referencial teórico adotado nesta pesquisa, a extração de informação (EI) ocorre com base na identificação de palavras-chaves em um conjunto de textos e na descrição dos padrões a elas associados, para, a partir desse ambiente lingüístico, retirar as informações buscadas.

Nesta análise, foi possível confirmar que, com efeito, cada palavra no contexto de ocorrência da palavra-chave participa ativamente do significado do núcleo e, conseqüentemente, contribui para o sentido do texto como um todo.

As relações léxico-semânticas estabelecidas pelas palavras-chaves são capazes de fornecer previsões para o leitor sobre algumas das possíveis interpretações para o texto, assim como levá-lo a levantar hipóteses durante o processo de leitura. De acordo com os autores citados neste trabalho (ECO, 1979; KLEIMAN, 2001, 2002; PEREIRA, 2003), o processo de levantamento de hipóteses e estabelecimento de previsões é essencial para o sucesso da interpretação de textos. Subsídios interessantes para esse processo podem ser encontrados a partir da observação dos padrões lexicais.

Por exemplo, retome-se o caso da palavra 'intestino'. A situação semântica em que esta palavra aparece como locativo (tanto como receptor quanto como órgão de passagem) leva o leitor a prever que o texto está abordando o funcionamento do processo digestivo. Aos poucos, o leitor confirma essa interpretação por meio da identificação dos itens lexicais associados à palavra-chave, extrai do texto as informações relativas a esse assunto e as organiza mais facilmente para a reconstrução mental do conteúdo do texto. Além disso, quando 'intestino' aparece como receptor de uma relação de causa e efeito, é possível que o leitor identifique

tópicos importantes: a patologia, alguns sintomas, ou, ainda, a localização dessas patologias no organismo.

Para o caso da palavra 'causa', a informação mais nítida está relacionada à identificação das patologias e seus sintomas, e essas informações ficam salientes a partir da ocorrência dos padrões lexicais. Nas relações lingüísticas demonstradas por estes padrões, o leitor pode identificar claramente a doença e seus efeitos no organismo, e é auxiliado a extrair essas informações, interpretando o texto com mais facilidade. Ainda com relação à palavra 'causa', evidenciou-se que o leitor a toma por base para construir lacunas interpretativas que serão mais facilmente preenchidas no decorrer da leitura do texto. Isto quer dizer que, ao identificar doença, algumas informações relacionadas aos seus sintomas e localização ficam em aberto para o leitor, que é auxiliado a preenchê-las por meio da observação de colocações que indicam essas informações, como, por exemplo, 'vômito', 'dor', 'queimação', etc.

Uma alta negatividade é averiguada tanto na palavra 'causa' quanto em '*Helicobacter Pylori*' (HP). As ocorrências desta última no *corpus* indicam, a exemplo das primeiras, o acontecimento de patologias. Embora 'HP' possua características semânticas diferentes das duas primeiras palavras analisadas, ela também facilita para o leitor a identificação das patologias e seus causadores. Por meio da observação do ambiente lingüístico em que aparece, é possível identificar, além dos traços semânticos desse microorganismo (bactéria, causador), os efeitos no organismo (úlcera, gastrite, inflamações gástricas).

Consideramos importante destacar que, ao deparar-se com todas essas palavras e ao organizar algumas informações, é possível retirar dos textos interpretações tais como:

- A digestão é um processo realizado em etapas, no qual vários órgãos são responsáveis. (papel semântico de passagem ou receptor de intestino).
- O intestino é responsável por uma parte da digestão e está exposto a doenças gástricas. (intestino como receptor de patologias gástricas).
- O exame da endoscopia é realizado para diagnosticar diferentes patologias gástricas. (relação de intestino com outros órgãos do sistema digestivo).
- A úlcera causa dor, queimação, desconforto. (relação causal na palavra 'causa').
- A gastrite e a úlcera, assim como inflamações gástricas podem ser causadas pela bactéria *Helicobacter Pylori*. (relação de HP como causador de patologias).

Essas seriam apenas algumas das muitas possibilidades de informações interpretativas que poderiam ser extraídas das relações léxico-semânticas mantidas pelas palavras-chaves selecionadas.

Destaca-se que, com efeito, essas seriam observações naturais tendo em vista o perfil dos textos de estudo. Contudo, percebe-se que, sendo esses os temas tratados nos textos, tais informações devem, obrigatoriamente, ser identificadas e retiradas pelo leitor para produzir interpretação. Reforçamos que tais informações foram retiradas mesmo sem acesso ao texto como um todo. A observação dos padrões facilitou a extração dessas informações. De fato, ao observar as ocorrências em contexto mais geral, foi possível constatar que as previsões e hipóteses formadas a partir das palavras-chaves são procedentes e estão de acordo com o conteúdo do texto.

Em suma, partimos do pressuposto de que para o leitor ser capaz de interpretar um texto, ele deve estar atento, entre outros aspectos, às relações mantidas pelas palavras e à maneira como elas se organizam para estabelecer sentido no texto. 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 ele 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 tornou-se evidente 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.

Deve-se destacar, no entanto, que não temos a intenção de extrair dos textos uma interpretação **geral (específica)** pela qual seja possível a representação formal da informação contida nos textos. O que pudemos averiguar aqui é que os padrões fornecem ao leitor uma possibilidade de facilitar o seu trabalho de interpretação, por intermédio da ativação de significados ocorrente na relação lingüística entre as palavras.

REFERÊNCIAS BIBLIOGRÁFICAS

ALLIENDE, Felipe; CONDEMARIN, Mabel. **A leitura: teoria, avaliação e desenvolvimento**. Porto Alegre: Artes Médicas, 1987.

ANTONINI, Eliana Pibernat. **Em busca do sujeito perdido**. Artigo apresentado em Congresso Brasileiro de Ciências da Comunicação, 2003. Disponível em: <http://intercom.locaweb.com.br/papers/congresso2003/nucleos_np01.shtml> Acesso em: abr. 2005.

BIBER, Douglas; CONRAD, Susan; REPPEN, Randi. **Corpus linguistics**: investigating language structure and use. Cambridge: Cambridge University Press, 1998.

CARDIE, Claire. A case based approach to knowledge acquisition for domain specific sentence. **Proceedings of the eleventh national conference on artificial intelligence**. AAAI press/ MIT press, 1993.

CAVALCANTI, Marilda do Couto. **Interação leitor-texto**: aspectos de interpretação pragmática. Campinas, SP: Editora da Unicamp, 1989.

COLOMER, Teresa; CAMPS, Anna. **Ensinar a ler, ensinar a compreender**. Tradução de Fátima Murad. Porto Alegre: Artmed, 2002.

COWIE, J. & WILKS, Y. Information extraction. In: DALE, R.; MOISL, H.; SOMERS, H. (eds.) **Handbook of natural language processing**. New York: Marcel Dekker, 2000. Disponível em: <<http://www.dcs.shef.ac.uk/~yorick/papers/infoext.pdf>> Acesso em: nov. 2004.

CRUSE, David A. **Lexical semantics**. Cambridge: Cambridge University Press, 1987.

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

ECO, Humberto. **Leitura do texto literário**. Tradução de Mario Brito. Lisboa: Editorial Presença, 1979.

HUFFMAN, Scott B. Learning information extraction patterns from examples. **Proceedings of the 1995 IJCAI workshop on new approaches to learning for natural language processing**. Abr. 1995. Disponível em: <http://ideas.repec.org/p/wop/prwawp/_039.html> Acesso em: out. 2004.

KATO, Mary A. **O aprendizado da leitura**. 5. ed. São Paulo: Martins Fontes, 1999.

_____. **No mundo da escrita**: uma perspectiva psicolingüística. 5. ed. São Paulo: Ática, 1995.

KLEIMAN, Ângela. **Oficina de leitura**: teoria e prática. 8. ed. Campinas, SP: Pontes, 2001.

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

_____. **Leitura**: ensino e pesquisa. 2. ed. Campinas, SP: Pontes, 2004.

LYONS, John. **Semântica**. Tradução de Wanda Ramos. Vol. I. Lisboa: Editorial Presença/Martins Fontes, 1977a.

_____. **Semantics**. Vol. II. Cambridge: Cambridge University Press, 1977b.

OLIVEIRA, Aparecida Augusta de. **Análise de textos da prova de inglês do vestibular da FUVEST**: um estudo baseado em corpus. Dissertação de Mestrado, PUC/SP, Ribeirão Preto: 2004.

ORLANDI, Eni Pulcinelli. **Interpretação**: autoria, leitura e efeitos do trabalho simbólico. Petrópolis: Vozes, 1996.

PARTINGTON, Alan. **Patterns and meanings**: using corpora for English language research and teaching. Amsterdam: John Benjamins Publishing Company, 1998.

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 RAM, Ashwin; MOORMAN, Kenneth (eds.). **Computational Models of Reading and Understanding**. Montreal: The MIT Press, 1999.

_____; WIEBE, Janyce. Learning extraction patterns for subjective expressions. **Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing**. (EMNLP-03).

_____; SCHMELZENBACH, Mark. An empirical approach to conceptual case frame acquisition. **Proceedings of the Sixth Workshop on Very Large Corpora**, 1998.

SARDINHA, Tony Berber. **Comparing corpora with wordsmith keywords**. São Paulo: LAEL/PUC-SP, 1996. Disponível em: <<http://acl.ldc.upenn.edu/W/W00/W00-0902.pdf>>. Acesso em: nov. 2003.

_____. **Estudo baseado em corpus da padronização lexical no português brasileiro: colocações e perfis semânticos**. São Paulo: LAEL/PUC-SP, 1999a. Disponível em: <http://www2.lael.pucsp.br/~tony/1999padroes_propor.pdf>. Acesso em: nov. 2003.

_____. **Looking at discourse in a corpus**: the role of lexical cohesion. São Paulo: LAEL/PUC-SP, 1999b. Disponível em: <<http://www2.lael.pucsp.br/~tony/2000aila.pdf>>. Acesso em: nov. 2003.

_____. Using keywords in text analysis: practical aspects. **Direct papers 42** São Paulo: LAEL/PUC-SP, 1999c. Disponível em: <<http://www2.lael.pucsp.br/direct/DirectPapers42.pdf>>. Acesso em: abr. 2004

_____. **Desafios da pesquisa assistida por computador na Lingüística Aplicada**: tecnologia e transdisciplinaridade. São Paulo: LAEL/PUC-SP, 2000a.

_____. **Os padrões de só**: um estudo baseado em corpus. São Paulo: LAEL/PUC-SP, 2000b. Disponível em: <http://www.lael.pucsp.br/~tony/2000so_propor.pdf>. Acesso em: abr. 2005.

_____. **Prosódia semântica na tradução do português e inglês**: um estudo baseado em corpus. São Paulo: LAEL/PUC-SP, 2000c

_____. Corpus Linguistics: history and problematization. **Delta**, vol.16, no.2., 2000d.

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

SINCLAIR, John. **Corpus, concordance, collocation**. Oxford: Oxford University Press, 1991. (Describing English Language Series).

SODERLAND, Stephen et al. Crystal: Inducing a conceptual dictionary. **Proceedings of the fourteenth international Joint conference on Artificial Intelligence**. 1995.

SPYRIDAKIS, Jan. Signaling effects: a review of the reseach – part I. **Journal of technical writing and communication**. University of Washington: Baywood Publishing. Vol.19 227-240, 1989a. Disponível em: <<http://www.uwtc.washington.edu/research/pubs/jspyridakis/>>. Acesso em: abr. 2005.

_____. Signaling effects: increased content retention and new answers – part II. **Journal of technical writing and communication**. Washington: Vol 19(4) 395-415, 1989b. Disponível em: <<http://www.uwtc.washington.edu/research/pubs/jspyridakis/>>. Acesso em: abr. 2005.

_____; ISAKSON, Carol S. The influence of semantics and syntax on what readers remember. **Journal of the society for technical communication**. Arlington, VA, USA. Vol. 50, number 4, 2003. Disponível em: <<http://www.uwtc.washington.edu/research/pubs/jspyridakis/>>. Acesso em: abr. 2005.

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

VERSPoor, Cornelia M., PAPEUN, George J., SENTZ, Kari. **A Theoretical motivation for patterns in Information extraction**. Los Alamos National Laboratory. Computer & Computational Science Division. Disponível em: <http://public.lanl.gov/verspoor/LAUR_03-1504.pdf>. Acesso em: maio 2004.

YANGARBER, Roman. & GRISHMAN, Ralph. **Extraction patterns discovery through corpus analysis**. New York University: Department of Computer Science. 2000.

