

ANÁLISE DE ASSUNTO E LINGÜÍSTICA TEXTUAL: UMA TENTATIVA DE APROXIMAÇÃO

Regina Helena van der Laan, CRB-10/514*

RESUMO: Aborda aspectos da Análise de Assunto e de Lingüística Textual. Estabelece uma inter-relação entre o trabalho de indexação e interpretação das questões de busca formuladas pelo usuário e a Lingüística Textual. Descreve a Análise de Assunto como um processo que permeia todo o trabalho de recuperação da informação, desde a indexação até as questões de expressas pelos usuários na busca de informação. Coloca que Lingüística Textual tem por objeto analisar a superfície do texto e compreendê-lo enquanto texto. Mostra que os princípios de análise do texto utilizados pela Lingüística Textual são semelhantes às etapas de Análise de Assunto, conforme a norma NBR12676 da ABNT. Baseia-se nos sete padrões de textualidade de Beaugrande. Exemplifica, utilizando-se de pesquisa realizada por McQuire e Eastman, como a Lingüística Textual pode auxiliar no processo de indexação e na compreensão das questões de busca dos usuários.

Palavras-chave: análise de assunto; lingüística textual; indexação; recuperação da informação.

INTRODUÇÃO

O objetivo deste trabalho é estabelecer uma inter-relação entre o trabalho de indexação e interpretação das questões de busca formuladas pelo usuário e a Lingüística Textual.

A Análise de Assunto é um processo que permeia todo o trabalho de recuperação da informação. Esta análise é feita ao indexar um documento e estabelecer os conceitos que devem ser indexados. Ou, ainda, ao analisar as questões propostas pelos usuários, visando

* Professora Assistente do Departamento de Ciências da Informação da UFRGS. Mestre em Educação pela PUC/RS. Doutoranda em Letras pela UFRGS. E-mail: vanderla@adufrgs.ufrgs.br

compreendê-las e traduzi-las para a linguagem utilizada pelos sistemas, a fim de recuperar o documento que contém a informação necessária ao usuário.

Neste processo estamos lidando com estruturas lingüísticas, que podem ser mal interpretadas pelo indexador levando-o a indexar um tópico diferentemente das intenções do autor.

No momento de interpretar as questões de busca dos usuários, a observação dos aspectos lingüísticos poderá auxiliar na determinação da estratégia de busca, o que permitirá uma resposta adequada às necessidades informacionais dos mesmos.

Iremos, inicialmente, focalizar alguns aspectos sobre Análise de Assunto. Num segundo momento, apresentaremos uma abordagem teórica sobre Lingüística Textual concluindo com os pontos de intercessão entre as duas áreas.

ANÁLISE DE ASSUNTO

O trabalho do bibliotecário tem como finalidade principal à recuperação de informações contidas em documentos. Para atingir esta finalidade, é necessário um criterioso trabalho de análise dos documentos para extrair os assuntos que deverão ser indexados.

O bibliotecário estabelece a interface entre a linguagem natural utilizada pelo autor e pelos usuários, em suas buscas de informação, e a linguagem artificial de indexação utilizada pelo Sistema de Recuperação de Informação. Segundo Cunha (1987, p.40) “. . . no momento da Análise Documentária existem três linguagens (para fins documentários, não lingüísticos) interligadas: a linguagem natural (LN),

as linguagens especializadas (LE) e as linguagens documentárias (LD).”

O processo de indexação está dividido em três estágios segundo a norma NBR12676 da ABNT:

- a) exame do documento e estabelecimento do assunto de seu conteúdo;
- b) identificação dos conceitos presentes no assunto;
- c) tradução desses conceitos para os termos de uma linguagem de indexação.

Análise segundo Barité (1997, p.16), é:

“Primeira fase del acto de clasificación, por la cual se realiza la lectura técnica, el análisis de contenido y la verificación de la forma del documento que va a ser clasificado. El análisis primario determina los asuntos o tópicos contenidos en el texto, en lenguaje natural. El clasificador o indizador registra las nociones, las sentencias o los términos más significativos, en el lenguaje propio del autor o en del analista.”

No primeiro estágio, a análise do documento, deve-se verificar inicialmente a forma física, pois o tipo de suporte irá determinar a estratégia a ser utilizada para análise.

Considerando a inviabilidade de uma leitura na íntegra do documento, estabelecem-se algumas estratégias para possibilitar a apropriação do conteúdo por parte do indexador. Esta leitura técnica deve considerar:

- a) o autor do texto, sua intenção. Deve-se perguntar com que objetivo o autor está comunicando este dado. A especialidade do autor, pois

isto irá permitir o direcionamento para o provável assunto a ser indexado;

- b) o título e subtítulo podem ajudar, mas jamais devem ser as únicas fontes de análise, uma vez que muitos títulos possuem uma grande ambigüidade que pode levar a uma interpretação equivocada do assunto tratado;
- c) resumo - esta leitura pode oferecer muitos subsídios, pois geralmente os resumos são escritos pelos próprios autores que indicam seus propósitos, os métodos utilizados e prováveis conclusões;
- d) sumário - considerando que o mesmo é a relação de todos os tópicos ou capítulos da obra, ele se torna uma preciosa fonte de análise para indexar, pois se pode encontrar nos títulos dos capítulos informações sobre os assuntos tratados pelos autores;
- e) introdução - as páginas introdutórias devem ser lidas na sua totalidade, isto irá possibilitar a compreensão do documento e a intenção do autor;
- f) análise de palavras em destaque no texto, grafadas em tipos diferentes ou negritadas;
- g) análise das referências bibliográficas.

No segundo estágio, o indexador, através de uma abordagem lógica, irá selecionar os conceitos que devem ser posteriormente indexados. Esta escolha pode ser feita através do estabelecimento de categorias como; o fenômeno tratado; o processo; propriedades, entre outras.

No último estágio, será feita a tradução dos conceitos a serem indexados para a Linguagem de Indexação adotada no Sistema de Informação.

Neste processo evidencia-se a importância dos conhecimentos prévios do indexador, seu conhecimento de mundo, seu conhecimento da área específica que está sendo indexada, sua capacidade de fazer inter-relações com outros textos já indexados.

O processo de Análise de Assunto, realizado no ato de indexação, também é feito, de certa forma, para analisar as questões de busca dos usuários.

O bibliotecário deve saber interpretar as solicitações dos usuários que, muitas vezes, não expressam claramente o que desejam. Isto ocorre porque eles não têm uma idéia muita clara do que procuram, ou acreditam que o Sistema não tem condições de fornecer nada a mais do que está sendo pedido. Um outro fator importante no atendimento de referência é que a maioria dos usuários busca informações sobre um determinado assunto, mas muito poucos procuram por um determinado autor ou um título específico. Este dado é importante para estabelecermos a relação entre a Lingüística Textual e a interpretação das questões de busca, uma vez que estas são expressas utilizando estruturas lingüísticas.

Um exemplo de importância da interpretação das questões dos usuários é um estudo feito por McQuire e Eastman (1998), sobre a ambigüidade da negação em questões de busca de informação. Conforme os autores “. . . the use of negation presents potentially serious problems and is not well handled by most systems.” (ibid., p.686).

Os autores propuseram para 64 estudantes de diferentes classes questões contendo o advérbio “não”. A pesquisa envolveu estudantes de lingüísticas, estudantes de programação de computador e estudantes de graduação em ciência da computação. Foram propostas 20 questões de busca contendo o advérbio de negação, com as prováveis interpretações. Somente em uma questão todos fizeram a mesma interpretação. Nas questões que obtiveram um percentual maior de coincidência nas interpretações, os itens negados estavam conectados pela conjunção “e”. Como na questão: “Books by Hawthorne but not Poe and Faulkner.” (ibid., 688) Este tipo de formulação foi categorizado como questão simples, de pouca dificuldade de entendimento.

Houve um grupo de onze questões que foram consideradas muito complexas. Muitas envolviam questões iniciadas com *sobre, por, antes, desde*.

Questões como :

Gostaria de obras sobre processo de linguagem natural mas não sobre lingüística, gramática e sintaxe.

poderiam gerar interpretações tais como:

- a) somente linguagem natural;
- b) linguagem natural e gramática;
- c) linguagem natural gramática e sintaxe;
- d) linguagem natural e sintaxe;

Verificou-se não estar claro se a negação referia-se à lingüística, gramática e/ou sintaxe. Estas ambigüidades geram um sério problema no momento do estabelecimento da estratégia de busca. A estratégia

estabelecida pode gerar como resultado da busca um alto índice de revocação com um baixo índice de precisão, sendo recuperados documentos não pertinentes aos interesses do pesquisador, gerando excesso de ruído. Ou, então, ao contrário, documentos importantes não são recuperados, gerando um constrangedor silêncio. Em ambos os casos, não iremos atender satisfatoriamente as necessidades reais de informação do leitor. Devido ao não entendimento dos aspectos lingüísticos o trabalho do profissional não será eficiente e muito menos eficaz.

Estas colocações confirmam nossa hipótese de que o conhecimento da teoria sobre Lingüística Textual pode tornar-se uma importante ferramenta no trabalho do bibliotecário, capacitando-o a compreender melhor os aspectos lingüísticos do texto e das questões de busca.

LINGÜÍSTICA TEXTUAL

A Lingüística Textual tem por objeto analisar a superfície do texto e compreendê-lo enquanto texto.

Segundo Ortiz (1998), a Lingüística Textual analisa o texto como:

- a) como imagem - analisa-se o título, o autor; a apresentação gráfica; formatos de parágrafo; a tipologia de impressão; formas de destaques; pontuação; numerações; ilustrações; etc.
- b) como mensagem - analisam-se as condições de produção e recepção. Neste tópico vamos analisar quem é o autor, quem é o destinatário do texto, local, data, meio de difusão; este dado é importante pois, se o texto está sendo divulgado em uma revista científica, terá uma linguagem e um público alvo específico; já um

texto escrito para um jornal ou revista de âmbito popular trará uma outra linguagem e teremos um outro tipo de leitor.

- c) como texto - nesta etapa será analisado o texto como tal observando-se nível enunciativo, nível narrativo, nível lógico e estratégias argumentativas.

Beaugrande (1981), afirma que o texto é uma OCORRÊNCIA COMUNICATIVA na qual encontramos sete padrões de TEXTUALIDADE:

1º padrão: COESÃO

Coesão é a maneira como os componentes se apresentam na superfície do texto, as palavras ouvidas ou lidas estão conectadas em uma seqüência. Os componentes do texto dependem das formas e convenções gramaticais. A coesão está ligada à dependência gramatical, ao sentido e uso. Os fatores de coesão permitem que um texto seja compreendido pelos falantes do mesmo idioma.

O autor afirma que a coesão de superfície textual e a subjacente coerência do mundo textual são os mais óbvios padrões de textualidade. Eles indicam como os elementos componentes do texto reúnem-se e fazem sentido.

Beaugrande (ibid.,1981), afirma ainda, que a interação entre coesão e outros padrões de textualidade fazem a comunicação eficiente.

2º padrão: COERÊNCIA

O autor continua afirmando que coerência é como os componentes do mundo textual configuram os conceitos e relações mutuamente acessíveis e relevantes. O conceito é definido como a configuração do conhecimento ao qual recorreremos mentalmente.

Relações são elos entre os conceitos que aparecem no mundo textual. Muitas vezes as relações não estão explícitas no texto. As pessoas constroem o sentido elaborando relações fora do texto.

Uma ação pode ser causa para a subsequente; pode possibilitar ou ser a razão da ação.

Outro ponto a salientar seria um arranjo temporal. Uma ação deve vir antes para causar a posterior.

Ao analisarmos o texto, podemos fazer inferências, deduzimos fatos que não estão explícitos no texto. Esta inferência é possibilitada através da relação do conhecimento do texto com o conhecimento do mundo.

3º padrão - INTENCIONALIDADE

A intencionalidade está relacionada ao produtor do texto, ao autor. Qual o seu propósito, qual o objetivo; qual seu plano. Considerando a intenção do autor, podemos verificar que o padrão de coesão e coerência irá atingir seu objetivo, tornando o texto plenamente compreensível. A configuração da linguagem deve ser intencional para a aceitação do texto como tal e de tal forma para ser utilizada na interação comunicativa. A produção e recepção dos textos funcionam como ações discursivas relevantes para alcançar algum propósito específico.

A inconsistência das estruturas de superfície, como quando o locutor muda o plano discursivo, emitindo uma declaração aparentemente sem sentido, pode ocasionar uma perda no entendimento e, portanto, não ocorrerá o processo de comunicação. Estas alterações discursivas são entendidas mais nas comunicações orais nas quais outros elementos de comunicação entram no processo

de compreensão da mensagem, tais como gestos, expressão corporal. Mas no texto escrito, objeto de nossa análise, isto pode ser altamente prejudicial, levando o interlocutor a não continuar a leitura por não ser capaz de compreender o texto lido. Na Análise de Assunto feita pelo bibliotecário isto se torna evidente, levando-o a indexar um texto, equivocadamente, em um assunto não tratado no mesmo.

4º padrão - ACEITABILIDADE

A aceitabilidade refere-se à atitude de quem recebe o texto. Envolve questões sociais, culturais. O produtor muitas vezes indaga como o leitor irá receber o texto. Cada indivíduo faz inserções e inferências conforme sua bagagem cultural. Mesmo o texto tendo coesão e coerência, ele tem que fazer sentido para o leitor.

5º padrão - INFORMATIVIDADE

É o grau de informação oferecido pelo autor. A obviedade e pouca informatividade podem resultar na não comunicação, e portanto na não aceitação do texto por parte do leitor. Este dado deve ser objeto de análise por parte do bibliotecário, pois uma informação pouco consistente pode gerar mais poluição no sistema do que resultar em um dado importante para o leitor.

6º padrão - SITUACIONALIDADE

A situacionalidade é entendida como os fatores que fazem um texto relevante para a situação de ocorrência.

7º padrão - INTERTEXTUALIDADE

É a relação com outros conhecimentos, com outros textos. O sentido do texto só poderá ser apreendido pelo leitor, caso o mesmo possa relacioná-lo com outros contextos já vividos ou lidos.

ANÁLISE DE ASSUNTO E LINGÜÍSTICA TEXTUAL

A relação entre a Lingüística Textual com a Indexação tornou-se evidente no instante em que se passou a traduzir a linguagem natural para a linguagem documentária, surgindo a necessidade de compreender os mecanismos lógicos e semânticos envolvidos.

Segundo Navarro (1988, p.57), “o analista da documentação deve levar em consideração algumas variáveis, principalmente as do discurso e deve ter em mente que essas variáveis não estarão presentes apenas no momento da análise, mas também na fase de obtenção do produto documentário.”

A relação que se estabelece entre a Análise de Assunto e a Lingüística Textual torna-se evidente no momento em que colocamos as etapas de análise, determinadas pela norma NBR12676 da ABNT, e comparamos com os tópicos abordados por Ortiz (1998).

Tanto a Análise de Assunto como a Lingüística Textual, num primeiro momento, analisam o texto sob aspectos gráficos e tipográficos, itens de apresentação, título, autoria, etc. Numa segunda etapa, outros aspectos devem ser observados como credenciais do autor, para quem foi escrito; isto para o indexador é fundamental, pois determinará o nível de linguagem que será utilizado na indexação. Se o texto é para um público leigo, os termos usados serão de nível popular; se for para especialistas, os termos serão em linguagem científica.

Para compreendermos um texto em processo de indexação, é evidente que o autor escreveu de forma coesa e coerente, caso contrário, ele provavelmente não irá atingir seus objetivos, não conseguindo transmitir sua mensagem. Se o texto não for compreensível, não será de interesse para o Serviço de Informação e,

portanto, não será adquirido. Os outros padrões de textualidade de Beaugrande evidenciam-se em todos os momentos do trabalho de análise, pois se o indexador não tiver conhecimento de mundo, não for capaz de ver os aspectos intertextuais provavelmente não irá indexar adequadamente, trazendo prejuízos futuros aos leitores.

Considerando, então:

- que o objeto de trabalho do bibliotecário é a informação;
- que esta é, basicamente, registrada em textos escritos conforme o código lingüístico de uma língua;
- que este texto deve ser compreensível para um determinado grupo de falantes;
- que a língua, seu registro, suas regras, sua evolução são o objeto de estudo da Lingüística;
- que a Lingüística Textual estuda os aspectos da superfície do texto;
- que a Análise de Assunto é o processo pelo qual o indexador apropria-se das idéias do autor e as sintetiza em forma de expressões lingüísticas usadas no vocabulário de indexação;
- que tanto os textos, como as questões de buscas formuladas pelos usuários estão carregadas de marcas lingüísticas, e que estas são objeto de estudo da Lingüística Textual,

verifica-se a relação entre as duas áreas de conhecimento e confirma-se a nossa hipótese inicial: que o conhecimento dos aspectos lingüísticos, em especial da Lingüística Textual, pode tornar-se um elemento de aperfeiçoamento no trabalho do bibliotecário.

SUBJECT ANALYSIS AND TEXTUAL LINGUISTICS: AN APPROACH

ABSTRACT: Deals with aspects of subject analysis and Textual Linguistics. Establishes an interrelation between the task of indexing and interpretation of search questions, posed by the user, and Textual Linguistics. Subject Analysis is described as a process that permeates all the work of information retrieval, from indexing to the questions expressed by users in their search for information. States that the aim of Textual Linguistics is analysing the surface of the text and understanding it as such. Demonstrates that the principles of text analysis used by Textual Linguistics are similar to the stages in Subject Analysis, as is prescribed by the rule NBR12676 from the ABNT. Is based on the seven patterns of textuality established by Beaugrande. Provides examples, using the research conducted by McQuire and Eastman, of how Textual Linguistics can help in the indexation process and in understanding search questions of users.

REFERÊNCIAS BIBLIOGRÁFICAS

ABNT. Métodos para análise de documentos - seleção de termos de indexação - NBR 12676. Rio de Janeiro, 1992.

BARITÉ, Mario Guido. Glosario sobre Organización y Representación del Conocimiento: Clasificación, Indización, Terminología. Montevideo, CSIC, Indice, 1997.

BEAUGRANDE, Robert-Alain de. Introduction to Text Linguistics. London; New York, Longman, 1981. p.1-13

CUNHA, Isabel M. R. Ferin. Análise Documentária. In: **Análise Documentária: a análise da síntese.** Coord. Johanna W, Smit. Brasília, IBICT, 1987.

MCQUIRE, April R.; EASTMAN, Caroline M. The Ambiguity of Negation in Natural Language Queries to Information Retrieval Systems. **Journal of the American Society for Information Science**, v.49, n.8, p.686-692, 1998.

NAVARRO, Sandrelei. Interface entre Lingüística e Indexação: revisão de literatura. **Revista Bras. de Bibliotec. e Doc.**, São Paulo, v.21, n.1/2, p.46-62, jan/jun. 1988.

ORTIZ, Elsa M. N. **Notas de aula**. Disciplina: Leituras Dirigidas: Tópicos de Lingüística Textual. Porto Alegre, PPG/Letras/UFRGS, 1998.